

**Computational Analysis of *Escherichia coli* O25 and O25b Carbohydrate  
Antigens using the CHARMM36 and GLYCAM06 Force Fields**

Alexander Rees Fourie  
Department of Computer Science  
University of Cape Town

In partial fulfillment of the requirements for the degree of MSc in Information Technology

February 2020

**Plagiarism Declaration**

I understand the meaning of plagiarism and declare that all of the work in this dissertation is either my own or has been properly acknowledged.

### **Acknowledgments**

Thank you to my research supervisor, Michelle Kuttel, for her support and guidance throughout this project and to the ICT High Performance Computing team at the University of Cape Town ([hpc.uct.ac.za](http://hpc.uct.ac.za)) for their help (and patience) in providing access to infrastructure for running simulations. This paper would not have been possible without you.

Thank you to my wife, Leana, for her support while I was working on this manuscript and for providing an ongoing supply of excellent coffee.

Finally, several key software applications made this study possible and are worthy of special mention: carbohydrate models for the CHARMM36 force field were constructed using v2 of the CarbBuilder application; carbohydrate models for the GLYCAM06 force field were constructed using the online GLYCAM-Web application; Molecular dynamics simulations were performed using the NAMD software application maintained by the Theoretical and Computational Biophysics Group at the University of Illinois; the Visual Molecular Dynamics (VMD) application proved invaluable in analysing and visualising simulation results.

### Abstract

The emergence of ST131 extra-intestinal pathogenic *Escherichia coli* that are resistant to multiple antibiotics is a growing international health concern. Infections are common, treatment options for antibiotic resistant bacteria are limited and there is no vaccine available.

Polysaccharides serve key functions in immune response to bacterial infection. The O-polysaccharides present on the cell surface of gram negative bacteria are antigenic and are associated with specific bacterial serogroups. These are, therefore, a potentially effective target for vaccines. Most ST131 *E. coli* isolates express the O25b antigen and monoclonal antibodies that are specific to it have been isolated.

The chemical structure of O25b has been characterized and differentiated from that of the previously known O25 (or O25a) variety. Relatively little is known about the conformations of O25a and O25b and how they differ, however. As conformation is a factor in antigen-antibody binding, differences between the conformations of these two antigens may be relevant to further research into carbohydrate targeted vaccines and diagnosis techniques for ST131:O25b bacteria.

The conformations of polysaccharides are typically dynamic in solution and are difficult to determine empirically. Molecular dynamics simulation provides a means of estimating polysaccharide conformation but the results are critically dependent on the quality of the selected force field. Carbohydrate force fields have matured over the past few decades and CHARMM36 and GLYCAM06 are used extensively for the analysis of bacterial polysaccharides. Studies that compare results from these two widely used force fields are, however, still quite rare.

Here we use molecular dynamics simulations of unacetylated, 3 RU oligosaccharide extensions to compare the CHARMM36 and GLYCAM06 force fields and to present an initial analysis of the conformations of the O25a and O25b *E. coli* antigens. We then apply CHARMM36 molecular dynamics simulation to analogous O- and N- acetylated oligosaccharide extensions to gauge the effect of these groups on the conformations of the two antigens and to compare O25a and O25b.

Despite some differences, our CHARMM36 and GLYCAM06 simulations are largely in agreement regarding the conformation of O25a trimers without O- or N-acetylation. Both force fields predict



extended, linear antigen conformations. Differences between the two force fields are noted in our analogous study of O25b however: GLYCAM06 favors a collapsed, globular oligosaccharide over a more extended molecule favored by CHARMM36; CHARMM36 and GLYCAM06 predict different preferred dihedral values for a conformationally important, main-chain  $\alpha$ -L-Rhap-(1->3)- $\beta$ -D-Glcp bond; GLYCAM06 favors an anti- $\Psi$ , anti- $\omega$  orientation of a side-chain  $\beta$ -D-Glc-(1->6)- $\alpha$ -D-Glc bond over an anti- $\Psi$ , syn- $\omega$  orientation favored by CHARMM36. These findings are in agreement with other studies that indicate the collapse of some oligosaccharides into metastable globular conformations during simulations with GLYCAM06.

Our CHARMM36 simulations of O- and N-acetylated, 3 RU oligosaccharide extensions of O25a and O25b indicate large differences between the conformations of the two antigens: First, the O25b trimer favors either a compressed or extended helical conformation in solution whereas the O25a trimer favors a single, extended conformation. Second, O25a and O25b exhibit notably different dihedral values for conformationally important glycosidic bonds that correspond with the reported structural differences between the two antigens. Third, O- and N-acetylation is found to facilitate rotation about a key  $\alpha$ -D-Glcp-(1->3)- $\alpha$ -L-Rhap2Ac bond in O25b that, in turn, facilitates the formation of compressed, helical O25b conformations. These compressed conformations are stabilized by intramolecular hydrogen bonds that involve O- and N-acetyl groups. Finally, N-acetyl groups appear to be shielded on the inside of the compressed O25b helix whereas O-acetyl groups appear to be exposed on the outside of the molecule. We postulate that these large conformational differences provide a rationale for the clinically noted differences in cross reactivity of monoclonal antibodies for the O25a and O25b antigens.

## Contents

Chapter 1: Introduction.....	8
1.1 Problem Statement.....	10
1.2 Aims.....	11
1.3 Research Questions.....	11
1.4 Approach.....	12
1.5 Overview.....	13
Chapter 2: Relevant carbohydrate chemistry.....	14
2.1 Monosaccharide stereo chemistry.....	14
2.2 Cyclisation of monosaccharides.....	14
2.4 O-Acetylation and N-acetylation.....	16
2.5 Glycosidic bonds, oligosaccharides and polysaccharides.....	16
Chapter 3: Computational modeling of oligo and polysaccharides.....	18
3.1 Molecular mechanics and force fields.....	20
3.2 Molecular dynamics.....	21
3.3 Metadynamics.....	21
Chapter 4: Antibiotic resistance and carbohydrate targeted vaccines.....	22
4.1 Antibiotic resistance.....	22
4.2 Carbohydrate targeted vaccines.....	23
Chapter 5: Escherichia coli serogroups O25 and O25b.....	24
5.1 Existing structural and conformational studies of <i>E. coli</i> O polysaccharides.....	26
Chapter 6: Methodology.....	29
6.1 Calculation of preferred disaccharide glycosidic bond dihedrals.....	29
6.2 Construction of aqueous phase 3 RU oligosaccharides models.....	30
6.3 Molecular dynamics simulations.....	30
6.4 Data analysis.....	31

Chapter 7: Results and discussion.....	34
7.1 CHARMM36 and GLYCAM06: Force field comparison.....	34
7.1.1 Preferred level of extension of O25a' and O25b' trimers.....	34
7.1.2 Preferred glycosidic bond dihedrals for O25a' and O25b'.....	40
7.1.3 RMSD cluster analysis of O25a' and O25b' trimers.....	49
7.1.4 Concluding remarks.....	52
7.2 Comparison of O25a and O25b conformations.....	52
7.2.1 Preferred level of extension of O25a and O25b trimers.....	53
7.2.2 Preferred glycosidic bond dihedrals for O25a and O25b.....	57
7.2.3 RMSD cluster analysis of O25a and O25b trimers.....	65
7.2.4 Concluding remarks.....	68
Chapter 8: Conclusions and future work.....	69
Chapter 9: References.....	71
Appendix A: Analysis of PMF contour plots of disaccharide fragments.....	79
Appendix B: Example NAMD configuration file for metadynamics simulations.....	83
Appendix C: Example metadynamics colvars definition file for disaccharide fragments.....	84
Appendix D: Example Gnuplot script for vacuum PMF surfaces.....	85
Appendix E: Example NAMD configuration file for solution phase simulations.....	86
Appendix F: Time series plot of r of O25a' and O25b' with selected conformations using the CHARMM36 and GLYCAM06 force fields.....	88
Appendix G: Time series and probability density plots of r of O25a', O25b', O25a and O25b for 200 ns of simulation with CHARMM36 and GLYCAM06.....	90
Appendix H: Extract of VMD Hydrogen bond analysis of O25b.....	92
Appendix I: Extract of VMD Hydrogen bond analysis of O25a.....	93

## Chapter 1: Introduction

Polysaccharides are key information carriers in biological processes including cell-cell interaction and host-pathogen recognition. The three dimensional structures, or conformations, of polysaccharides are integral to their biological activity and their elucidation is a central problem in glycobiology.<sup>1-4</sup> Monosaccharides contain chiral centers, present reactive hydroxyl groups at several positions and are interlinked by flexible glycosidic bonds in polysaccharides.<sup>5</sup> Due to the flexible nature of glycosidic bonds, polysaccharides adopt a wide variety of conformations that are typically dynamic in aqueous solution and are, therefore, not adequately represented by static models.<sup>6</sup>

The emergence of pathogenic multi-drug resistant gram-negative bacteria is a growing international health concern. Although antibiotics are often still effective, the pipeline for the development of new treatments for infection may not be sufficient to keep up with the emergence of resistant organisms.<sup>7</sup> This has stimulated renewed interest in the development of bacterial surface carbohydrate targeted vaccines and diagnosis techniques to prevent and to quickly identify infections.<sup>8</sup>

Biologically active conjugates of lipids and polysaccharides (lipopolysaccharides) are present on the cell surface of gram-negative bacteria. Lipopolysaccharide structure varies between bacterial species but consists of lipid A, core oligosaccharide and O-polysaccharide regions. The lipid A region is embedded in the cell membrane and is attached to the core oligosaccharide. The O-polysaccharide is attached to the core, is made up of a series of repeating monosaccharide units (RUs), is variable and is the determinant of bacterial serogroup.<sup>8</sup> O-polysaccharide is antigenic and presents an accessible target for the development of serogroup specific vaccines and diagnosis techniques.<sup>9,10</sup> The variability of O-polysaccharide within bacterial species, however, means that it is usually only economical to target it in scenarios where a few serogroups of a pathogen are the most prevalent.<sup>11</sup>

Molecular mechanics and molecular dynamics are computer simulation methods for modeling physical systems ranging from simple fluids to complex materials. These methods have been evolving since their inception in the 1950s and have been bolstered by improvements to computing facilities and their availability to the scientific community.<sup>12,13</sup> Modeling and molecular dynamics

simulations of polysaccharides provides insight into their conformations and interactions with other molecules that are very difficult to glean using experimental data alone.

*Escherichia coli* is a well characterized, common gram-negative bacterium.<sup>14,15</sup> Not all *E. coli* cause disease<sup>16</sup> but those from the pathogenic ST131-O25b:H4 lineage cause community and hospital acquired multi-drug resistant infections. Cases occur internationally with the spectrum of disease ranging from cystitis to life-threatening sepsis.<sup>17,18</sup> ST131-O25b:H4 includes most fluoroquinolone resistant *E. coli* and about half of the isolates that express extended-spectrum  $\beta$ -lactamases (that provide resistance to  $\beta$ -lactam antibiotics).<sup>10,19</sup> Reports of representative strains that express carbapenemases have begun to appear and are concerning as treatment options for these infections are very limited.<sup>19</sup>

The high occurrence of infection and multi-drug resistance of many *E. coli* ST131-O25b:H4 isolates make the cell surface lipopolysaccharides that they express an interesting target for investigation. An understanding of the conformation of the O25b polysaccharide and how it differs from that of the classical O25 (or O25a) variety may contribute to the development of serogroup specific vaccines and diagnosis techniques.<sup>20</sup>

Empirical investigations into the structures of the repeating units (RUs) of O25a and O25b and have revealed a difference in their central monosaccharides:<sup>10,14</sup>

O25a: [  $\rightarrow$  4)- $\alpha$ -D-Glcp\*-(1-3)- **$\alpha$ -L-FucpNAc**-(1-3)- $\beta$ -D-GlcpNAc-(1  $\rightarrow$  ]

O25b: [  $\rightarrow$  4)- $\alpha$ -D-Glcp\*-(1-3)- **$\alpha$ -L-RhapAc**-(1-3)- $\beta$ -D-GlcpNAc-(1  $\rightarrow$  ]

\* 3, 6 substitution with  $\alpha$ -L-Rhap and  $\beta$ -D-Glcp

Szijarto *et. al*<sup>9,10</sup> report that “Immunoblot staining with commercially available O25 typing serum has confirmed some cross-reactivity between the O25a and O25b antigens but with much lower reactivity to O25b. Monoclonal antibodies that bind to the surface of live *E. coli* O25b strains irrespective of the capsular type expressed, and that do not bind to bacteria or purified lipopolysaccharide from other serotypes, including the O25a antigen, have been identified.”

Computational studies of other bacterial polysaccharides have found that structural differences in their RUs may impact polysaccharide conformation and hence have an impact on the cross

reactivity of antibodies.<sup>2,3,21</sup> As such, molecular dynamics simulations of O25a and O25b may serve to highlight such differences and provide a conformational rationale for clinical observations.

A force field is a mathematical model that relates the forces acting on the atoms of a molecular system to the potential energy of the system over time.<sup>13</sup> Force fields are comprised of a potential energy function and an associated set of parameters for a particular class of molecules. The selection of an appropriate force field is a key consideration for molecular dynamics simulation of carbohydrates and has a significant effect on the outcome.<sup>2,22</sup> Carbohydrate force fields have been enhanced over the past few decades<sup>22,23</sup> and the CHARMM36<sup>24-28</sup> and GLYCAM06<sup>29</sup> force fields are now applied extensively to the study of bacterial surface carbohydrates.

Duplicate simulations using both CHARMM36 and GLYCAM06 affords the opportunity to gain insights into the differences in predictions that may occur due to force field selection. Such results are valuable as studies that compare simulation results for carbohydrates using more than one force field are relatively rare and, where they have been undertaken, have revealed important differences for some molecules.<sup>2,22</sup>

## 1.1 Problem Statement

There is no prophylactic vaccine against infection with multi-drug resistant ST131-O25b:H4 *E. coli* available. The O25b polysaccharide is an accessible bacterial antigen that is structurally different from the previously known O25a and is a target for monoclonal antibodies that do not cross-react with O25a. Conformational differences between O25a and O25b may provide a rationale for these (and other) clinical observations but carbohydrate conformations are difficult to glean from experimental data and few (if any) computational studies of O25a and O25b are available.

The selection of the most appropriate force field for a molecular dynamics simulation is key to obtaining realistic results. The CHARMM36 and GLYCAM06 force fields are used extensively for carbohydrate simulations but studies that compare results are relatively rare. Modeling results are critically dependent on the selected force field so it is important to know if there are significant differences between the two force fields most commonly used to model carbohydrates.

## 1.2 Aims

This study will provide insight into the preferred conformations of the O25a and O25b polysaccharides via molecular dynamics simulation using the CHARMM36 carbohydrate force field. 3 RU O25a and O25b oligosaccharides, with and without acetylation, will be modeled in aqueous solution. The preferred end-to-end distances of the oligosaccharides will be assessed and the preferred glycosidic bond dihedrals of their central RUs will be calculated. Finally, the effect of O- and N-acetyl side chains on predicted glycosidic bond dihedrals and on the end-to-end distances of the oligosaccharides will be determined.

In addition, this work will compare molecular dynamics simulation results from the CHARMM36 and GLYCAM06 carbohydrate force fields. Models of 3 RU O25a and O25b oligosaccharides without O- and N-acetyl groups will be used as a case study for this purpose. The preferred levels of extension of the two oligosaccharides and the predicted glycosidic bond dihedrals of their central RUs will be assessed using both force fields and the results will be compared.

## 1.3 Research Questions

1. **What effect does the established structural difference between the repeating units of O25a and O25b have on their conformations?** Differences in antigen confirmation may have an effect on antigen-antibody binding.<sup>20,22</sup> So, how different are the preferred conformations of the O25a and O25b polysaccharides?
2. **What is the effect of O- and N-acetylation on the conformations of O25a and O25b?** According to Kuttel *et. al.*<sup>22</sup>, “O-acetylation is a common modification of bacterial polysaccharides but its impact on conformation is poorly understood and polysaccharide dependent.” So, does O- and N-acetylation of monosaccharides have a significant effect on the conformations of the O25a and O25b polysaccharides?
3. **Is there a significant difference in the conformations predicted by the CHARMM36 and GLYCAM06 carbohydrate force fields?** Previous studies indicate that somewhat different molecular dynamics simulation results can be expected.<sup>2,22</sup> So, do CHARMM36 and GLYCAM06 predict different conformations for the O25a and O25b polysaccharides?

## 1.4 Approach

The systematic approach recommended by M. Kuttel and N. Ravenscroft<sup>22</sup> is applied to the analysis of the O25a and O25b polysaccharide conformations here. Static models of twelve disaccharides representing the glycosidic bonds of the O25a and O25b RUs (with and without O-acetyl and N-acetyl groups) are constructed. Each disaccharide is subjected to biased (Metadynamics) molecular dynamics simulation<sup>30,31</sup> in vacuum using their glycosidic bond dihedrals as collective variables. Disaccharides with O- or N-acetyl groups are simulated using only the CHARMM36 force field. All other disaccharides are simulated in duplicate using both the CHARMM36 and GLYCAM06 force fields. A Potential of Mean Force (PMF) surface is calculated from the results of each simulation and used to find low energy dihedrals for each glycosidic bond.

Four 3 RU static oligosaccharide extensions of O25a and O25b (with and without O- or N-acetyl groups) are constructed using dihedral values informed by the low energy regions of the PMF surfaces. Each oligosaccharide is immersed in aqueous solution and is subjected to unbiased molecular dynamics simulation using periodic boundary conditions. Oligosaccharide extensions with O-acetyl or N-acetyl groups are only simulated with the CHARMM36 force field. All other oligosaccharide extensions are simulated in duplicate using both the CHARMM36 and GLYCAM06 force fields.

Time series dihedral values for each glycosidic bond are extracted from the unbiased simulation trajectories and used to visualize the conformation space explored over the simulation period. Time series distance values between selected atoms on the first and last RUs are used to visualize the preferred end-to-end distance of each oligosaccharide. Comparisons of the data gleaned from CHARMM36 and GLYCAM06 simulations are drawn where simulations were performed with both force fields (i.e. for oligosaccharides without O-acetyl or N-acetyl groups). As CHARMM simulations of 3 RU oligosaccharides that both include and exclude O-acetyl and N-acetyl groups were performed, the effects of these groups on oligosaccharide conformation can be assessed by comparing these results.



## 1.5 Overview

Chapter 2 introduces some carbohydrate chemistry that is relevant to the rest of the text. The focus is on describing the molecular structures and stereo chemistry of mono, oligo and polysaccharides as the basis for constructing computational models.

Chapter 3 provides an introduction to computational modeling of glycosidic bonds and, by extension, oligo and polysaccharides. Molecular mechanics force fields, molecular dynamics simulation and Metadynamics are introduced.

Chapter 4 presents a brief discussion of the emergence of antibiotic resistance and introduces carbohydrate targeted vaccines. The structure of lipopolysaccharide present on the surface of gram negative bacteria is described and O-polysaccharide is positioned as an antigen.

Chapter 5 is focused on the O25a and O25b serogroups of *E. coli*. A discussion of the emergence of the multi-drug resistance ST131-O25b:H4 lineage is presented and the molecular structures of the O25a and O25b repeating units are described.

Chapter 6 covers the methodology applied in this study. Modeling of oligosaccharide extensions, configuration information for molecular dynamics simulations and our approach to data analysis are described in detail here.

Chapter 7 presents the results and our discussion of the modeling and molecular dynamics simulations of the O25a and O25b polysaccharides and their representative disaccharide fragments.

Chapter 8 draws conclusions from the work performed in this study and includes some recommendations for future work.

Chapter 9 provides references for other work cited in the rest of the text.

## Chapter 2: Relevant carbohydrate chemistry

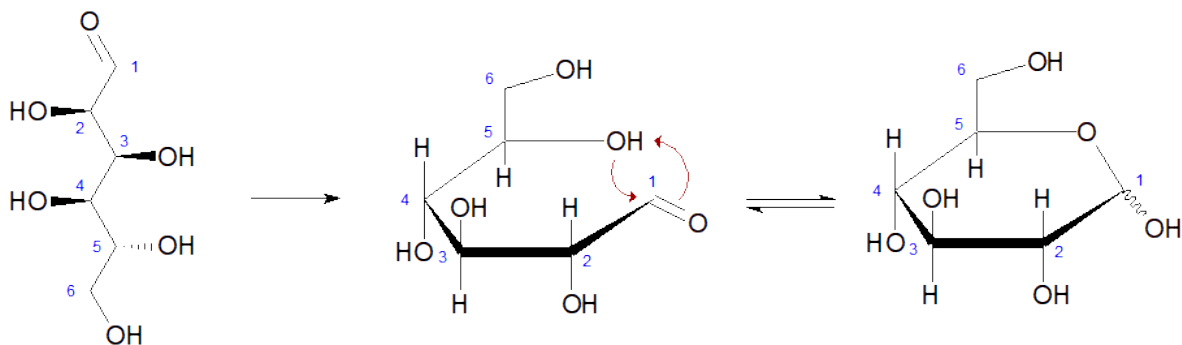
Carbohydrates are often defined as “polyhydroxy aldehydes or ketones or substances that yield these compounds on hydrolysis.”<sup>5</sup> Monosaccharides are the simplest carbohydrates and consist of a single polyhydroxy aldehyde or ketone unit of variable carbon chain length. Monosaccharides are classified as either aldoses or ketoses based on the position of their carbonyl group: those with a carbonyl group at the end of their carbon chains are classed as aldoses and those with a carbonyl group positioned elsewhere are classed as ketoses.<sup>5</sup> Monosaccharides may be comprised of chains of between five and seven carbon atoms. Those made up of chains of six carbon atoms are referred to as hexoses<sup>5</sup>, are common in bacterial lipopolysaccharides<sup>32</sup> and will be the focus of this study.

### 2.1 Monosaccharide stereo chemistry

Each carbon atom of a monosaccharide, except for the two terminal carbon atoms, constitutes a chiral center so that there are  $2^4$  possible stereo isomers of a hexose. By convention, the stereo isomers of a monosaccharide are classified as either D or L based on the configuration of the chiral center furthest from the carbonyl group. A stereo isomer is designated D if its configuration is the same as that of the only chiral center of D-Glyceraldehyde and L if it is the same as that of L-Glyceraldehyde. Monosaccharides that differ only in the configuration of a single chiral center are known as epimers.<sup>5</sup>

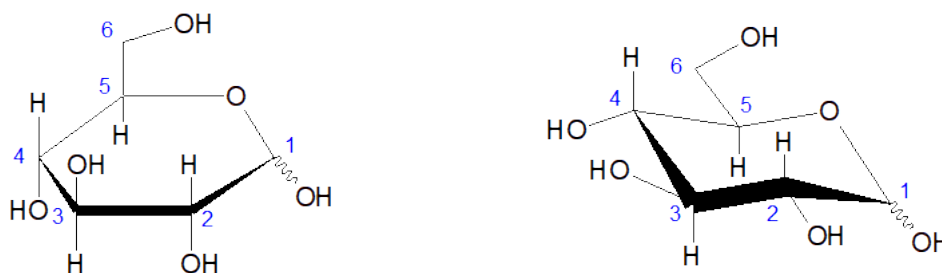
### 2.2 Cyclisation of monosaccharides

Hexoses exist almost entirely as cyclic hemiacetals, referred to as pyranoses, in solution. Pyranoses are formed by intramolecular attack of the hydroxyl group of the carbon 5 atom (C5) on the terminal carbonyl group of the C1 atom as illustrated in figure 1.



*Figure 1: Stereo (left) and Haworth (middle and right) formulae of the cyclisation of D-Glucose into D-Glucopyranose. Carbon atoms are labeled C1 through to C6 by convention.<sup>5</sup>*

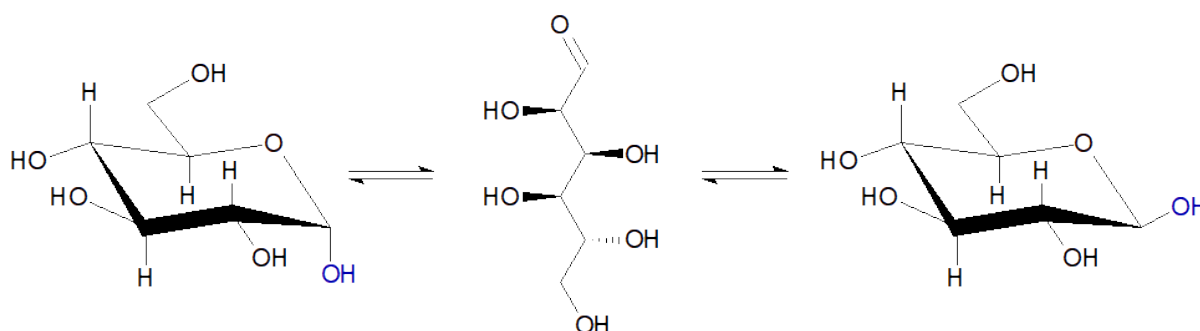
Following cyclisation, pyranoses adopt a fairly rigid, puckered ring conformation in solution. This arrangement (the “chair conformation”) maximizes the distance between the C1 and C4 atoms, has a carbon-carbon bond angle of close to 109.5 degrees (optimal for  $sp^3$  orbitals) and is largely free from torsional ring strain. So called “boat” and “twisted” pyranose ring configurations are possible but are typically more strained and are not considered further in this study.<sup>5</sup>



*Figure 2: Molecular representations of D-Glucopyranose depicting the puckered ring structure adopted by pyranoses in solution and the positions of hydroxyl groups above and below the ring.*

Pyranoses have a chiral center at C1 which is referred to as the anomeric carbon. Two stereo isomers, known as alpha ( $\alpha$ ) and beta ( $\beta$ ) anomers, that differ only in the orientation of the anomeric group are possible.<sup>5</sup> The group is in an axial position (almost perpendicular to the plane of the ring) in the alpha anomer and in an equatorial position in the beta anomer, as illustrated in figure 3.

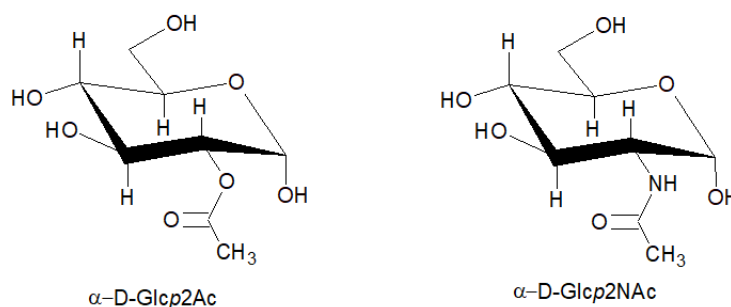
The concentration of straight chain hexoses in aqueous solution is negligible and an equilibrium exists between  $\alpha$  and  $\beta$  anomers. The equilibrium concentrations of anomers of different monosaccharides vary due to the anomeric effect<sup>33</sup>, steric effects between the ring groups (such as hydroxyl groups or attached O-acetyl or N-acetyl groups) and interactions with solvents.<sup>5</sup>



*Figure 3: Mutarotation results in an equilibrium mixture of alpha (left) and beta (right) anomers of pyranoses in solution.*

## 2.4 O-Acetylation and N-acetylation

Monosaccharides that have O-acetylated or N-acetylated hydroxyl groups may be present in bacterial lipopolysaccharides and, indeed, both the *E. coli* O25a and O25b polysaccharides contain acetylated or N-acetylated units.<sup>9,16</sup> O-Acetylation and N-acetylation may affect how monosaccharides will be orientated when they are bridged by glycosidic bonds and ultimately may have an impact on the conformations adopted by oligo- and polysaccharides.<sup>22</sup>

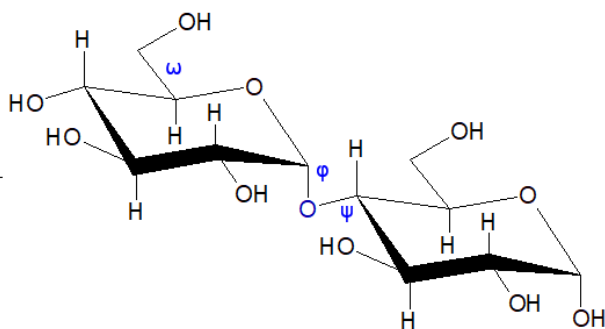


*Figure 4: Molecular representations of O-acetylated (left) and N-acetylated (right) D-Glucopyranose including abbreviated notations for each molecule.*

## 2.5 Glycosidic bonds, oligosaccharides and polysaccharides

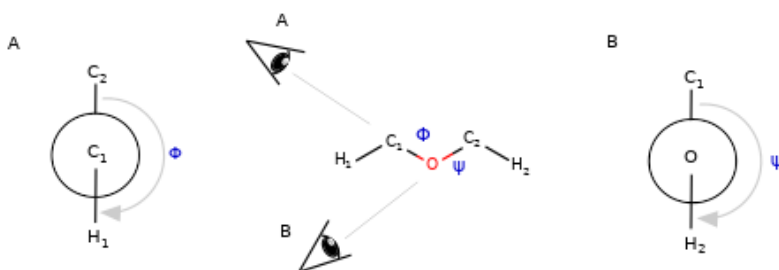
Disaccharides, trisaccharides, oligo- and polysaccharides are comprised of two, three or more monosaccharide units linked by glycosidic bonds.<sup>5</sup> Glycosidic bonds between monosaccharides form by the elimination of water from the anomeric hydroxyl group of one monosaccharide and a hydroxyl group of another. Bonds may form from hydroxyl groups at various positions resulting in different types of glycosidic linkage. (C1-C4), (C1-C3) and (C1-C6) linkages are present in the *E. coli* O25a and O25b RUs<sup>9,16</sup> but other linkages are also possible. Oligosaccharides are typically comprised of a few linked monosaccharides (not more than 20) whereas carbohydrates containing more units are referred to as polysaccharides.

As discussed in Section 2.2, pyranose rings are fairly rigid. The glycosidic bonds that bridge monosaccharides to form polymers are, however, much more flexible. The orientations of pyranose rings around glycosidic bonds are thus the primary determinant of the conformations of oligo and polysaccharides. The dihedral angles phi ( $\phi$ ), psi ( $\psi$ ) and omega ( $\omega$ ), illustrated in Figure 5, describe the orientation of rings about an example (C1-C4) glycosidic bond.<sup>34</sup>



*Figure 5: Maltose comprised of two  $\alpha$ -D-Glucopyranose monosaccharides linked by a C1-C4 glycosidic bond. Glycosidic bond dihedrals  $\phi$ ,  $\psi$  and  $\omega$  are highlighted in blue.*

In this work, the dihedrals  $\phi$  and  $\psi$  are defined as the angles between the following sets of atoms on two pyranose rings adjoined by a glycosidic bond:  $\phi = \{H_1 - C_1 - O_x' - C_x'\}$ ;  $\psi = \{C_1 - O_x' - C_x' - H_x'\}$ .  $\omega$  is defined analogously to  $\phi$  and  $\psi$  for bond involving a primary hydroxyl group as  $\omega = \{O_6' - C_6' - C_5' - O_5'\}$ . A simplified diagram demonstrating values of  $\phi$  and  $\psi$  of 180 degrees is presented in figure 6. Arrangements of dihedral angles in the range of  $\{-90 < \text{Angle} < 90\}$  degrees are referred to as “syn” conformers whereas arrangements with angles in the range of  $\{-90 > \text{Angle} > -180\}$  are referred to as “anti” conformers.



*Figure 6: Dihedral angles  $\phi$  and  $\psi$  of 180 degrees illustrating an anti arrangements about the indicated  $(C_1 - O)$  and  $(O - C_2)$  bonds.*

As glycosidic bonds are flexible, their dihedral angles are not fixed but rather occupy a range of possible values with frequencies defined by some polyfactorial probability function. The conformations of glycosides are, therefore, dependent on all of the factors that influence this function, are dynamic and are not adequately represented by static models. Computational molecular modeling and molecular dynamics simulation provides a means of exploring the conformation space available to oligo and polysaccharides and is the subject of chapter 3.

### Chapter 3: Computational modeling of oligo and polysaccharides

Section 2.5 established that the set of dihedral angles,  $\phi$ ,  $\psi$  and  $\omega$ , are key criteria for describing the conformations of oligo and polysaccharides and, furthermore, that they occupy a range of possible values with frequencies defined by some probability function. The Boltzmann (or Gibbs) distribution is a probability function that describes the likelihood that a system will occur in a certain state in relation to the energy of the state and the temperature of the system.<sup>35</sup>

$$p_i = \frac{e^{-\epsilon_i/kT}}{\sum_{j=1}^M e^{-\epsilon_j/kT}} = c e^{-\epsilon_i/kT}$$

*Equation 1: The Boltzmann distribution for a state  $p_i$  with a thermodynamic temperature  $T$  and an energy  $\epsilon_i$  where  $M$  is the total number of possible states,  $k$  is Boltzmann's constant and  $c$  is a normalisation constant.*

If the relative probability of a system occurring in one of two states,  $p_i$  and  $p_j$ , at a constant temperature is considered then the normalisation constant in Equation 1 is not relevant:

$$\frac{p_i}{p_j} = e^{\frac{\epsilon_j - \epsilon_i}{kT}}$$

*Equation 2: The Boltzmann factor ( $p_i/p_j$ ) for two states  $i$  and  $j$ . The ratio of the probabilities of occurrence of the states depends on the energy difference ( $\epsilon_j - \epsilon_i$ ) between them such that the low energy state is favored.  $k$  is the Boltzmann constant and  $T$  is the temperature of the system.*

Oligo and polysaccharides that are of biological interest typically occur in aqueous solution under constant temperature and pressure. The energy values ( $\epsilon_j$  and  $\epsilon_i$ ) may be replaced by enthalpy for such systems:

$$H = U + pV \Rightarrow \frac{p_i}{p_j} = e^{\frac{H_j - H_i}{kT}} \quad \text{or} \quad p_i = c e^{-H_i/kT} \quad (\text{from Equation 1})$$

*Equation 3: Substitution of enthalpy,  $H$ , for energy in the Boltzmann distribution for systems under constant pressure and temperature.  $U$ ,  $p$  and  $V$  are the internal energy, pressure and volume of the system respectively.*

In accordance with the second law of thermodynamics, the entropy of a system is a function of the number of possible states available to it:

$$S = k \ln(N) \Rightarrow N = e^{S/k}$$

*Equation 4: The entropy,  $S$ , of a system with  $N$  possible states.  $k$  is Boltzmann's constant.*

Multiple arrangements of atoms in a chemical system with the same energy are possible and, as such, the Boltzmann factor is multiplied by the total number of possible states at a given energy to reflect the overall probability of occurrence of a state. Multiplication of the Boltzmann factor in Equation 3 by  $N$  in Equation 4 (with some simplification) yields:

$$p_i = c e^{-(H-TS)/kT} \quad \text{or} \quad p_i = c e^{-G/kT} \quad \text{as} \quad G = H - TS$$

*Equation 5: The relationship between Gibbs Free Energy,  $G$ , and the probability of occurrence of a state of a system at a constant temperature and pressure.*

The important relation presented in Equation 5 indicates that the probability of occurrence of a state,  $p_i$ , is proportional to its Gibbs Free Energy,  $G$ . From this relation, and the discussion presented in Section 2.5, it is apparent that the most probable values of  $\phi$ ,  $\psi$  and  $\omega$  for a glycosidic bond are those that minimize the free energy contribution of the bond to the overall energy of the system. A common approach to evaluating low energy dihedral values is to use computational methods to calculate the energy of a disaccharide in many iterations as the values of  $\phi$ ,  $\psi$  and  $\omega$  are varied. This data is suitable for visualisation in two dimensions and describes the relationship between the dihedral angles and  $G$  for a glycosidic bond.<sup>23</sup> This topic is discussed further in Sections 3.2 and 3.3 but it is necessary to digress here to consider how the overall energy of a molecule in a given conformation can be calculated.

Each dihedral angle constitutes a single coordinate dimension to a potential energy hypersurface for a molecule that is comprised of  $3N - 6$  dimensions for  $N > 2$  atoms. Potential energy contributions for bond lengths, bond angles, improper angles, 1-3 cross term angle bending (Urey-Bradley potential) and non-bonded interactions for each atom are also contributing factors. Molecular mechanics and so called "force fields" provide a means of calculating the overall potential energy surface of a molecule by evaluating energy contributions for each atom and are the subject of Section 3.1.

### 3.1 Molecular mechanics and force fields

According to Zhu *et. al* <sup>23</sup>, “Molecular mechanics treats atoms and covalent bonds using a balls-on-springs approach that allows molecules to be modeled using Newtonian mechanics. The properties of the system are defined by the elasticity of covalent bonds, valance and dihedral angles and electronic and Van der Waals interactions between non-bonded atoms.” An example potential energy function for the CHARMM additive force field is illustrated in Equation 6. <sup>23</sup>

$$U(R) = \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{UB} K_{UB} (S - S_0)^2 + \sum_{dihedrals} K_\chi (1 + \cos(n_\chi - \delta)) \\ + \sum_{impropers} K_{imp} (\phi - \phi_0)^2 + \sum_{nonbond} \left( \epsilon_{ij} \left[ \left( \frac{R_{(min_{ij})}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{(min_{ij})}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon_l r_{ij}} \right)$$

Equation 6: Pair-wise potential energy function of the CHARMM additive force field. <sup>23</sup>

A force field is comprised of a potential energy function and a set of parameters for a selected class of molecules, such as carbohydrates, that together provide realistic models of specific molecules. Force field parameters are defined through an iterative process using experimental or quantum mechanical data to optimize the input values of the energy function. <sup>22,23</sup> This study makes use of the existing CHARMM36 and GLYCAM06 carbohydrate force fields developed by Alex MacKerell Jr.’s research team at the University of Maryland <sup>24–28,36</sup> and Robert Woods’ research group at the University of Georgia respectively <sup>29</sup>.

Force fields for carbohydrates have matured significantly over the past decade. The CHARMM36 and GLYCAM06 force fields have been used extensively to model oligo and polysaccharides and include parameters for glycosidic linkages between most types of monosaccharide. <sup>22,23</sup> Parameters for modeling glycosidic linkages between some monosaccharides with O-acetyl and N-acetyl side groups are now available in these two force fields.

Software for generating and validating initial three dimensional structures of carbohydrates exist and include web-based tools such as Glycam-Web <sup>37</sup> (for building models for use with GLYCAM based molecular dynamics simulations) and desktop applications such as Carbbuilder <sup>1</sup> (for building models for use with CHARMM based simulations).



### 3.2 Molecular dynamics

Using a suitable force field and starting structure, molecular dynamics simulation calculates the motion, interactions and energy of a molecular system over time.<sup>23</sup> The velocity and forces acting on each atom are evaluated at discrete time intervals and recorded into a trajectory file that contains a large amount of information about how the system changes over time.

Molecular dynamics simulation in conjunction with Metadynamics<sup>30</sup> can be used to produce a Potential of Mean Force (PMF) surface which, according to French *et. al*<sup>38</sup>, “represents the change in free energy of a system as a function of an internal coordinate or external parameter.” If the angles  $\phi$  and  $\psi$  for a glycosidic linkage are selected as internal co-ordinates of interest then the energy of the system can be calculated as their values are varied. A plot of free energy against the values of the two angles yields a surface referred to as a PMF, energy map or Ramachandran plot. The global minimum of such a plot represents the lowest energy geometry of the selected coordinates.<sup>38</sup>

### 3.3 Metadynamics

As discussed previously, the internal geometry of pyranose rings is stable relative to glycosidic bonds that bridge monosaccharides. Steric interactions inhibit the free rotation of pyranose rings about glycosidic bonds in polysaccharides and may result in relatively high-energy intermediates for transitions from one relatively stable conformation to another. As a result, simulation of rotation about glycosidic bonds may result in local energy minima for some dihedral values that require long, computationally expensive molecular dynamics simulation times to escape.

The Metadynamics routine described by Laio *et. al.*<sup>30</sup> provides a means of escaping local energy minima by progressively applying small repulsive Gaussian potentials to explored regions of the energy surface as a simulation progresses. Metadynamics is well suited to the simulation of rare molecular events for two or three collective variables<sup>30</sup> such as rotation about  $\psi$ ,  $\phi$  and  $\omega$  dihedral angles.

## Chapter 4: Antibiotic resistance and carbohydrate targeted vaccines

### 4.1 Antibiotic resistance

Prior to the discovery and application of antibiotics (such as penicillin) in the 1940s, infectious diseases were a leading cause of death worldwide. Clinicians were optimistic about the effectiveness of these newly discovered anti-microbial agents and applied them liberally in the treatment of a wide range of infections. The capability of some bacteria to develop resistance to antibiotics was noted early (around 1929) but, due to their overwhelming success in treating infections, these findings were not given much attention. Several isolates of staphylococci that were resistant to penicillin were reported about a year after the introduction of the drug and, by 1947, hospital acquired infections with *Staphylococcus aureus* were usually resistant to penicillin G and multiple other types of antibiotics. Isolates of *E. coli* (and other gram-negative bacteria) that exhibited resistance to antibiotics were identified and characterized in the 1960s. The chemical modification of antibiotic functional groups during the 1970s and 1980s yielded a wide range of variants that were initially effective but there has been a steady increase in the emergence of multi-drug resistant bacteria since then.<sup>39</sup> This has been attributed to the general overuse of antibiotics, poor infection control in and around hospitals and the overuse of antibiotics in animal feeds and for the treatment of animals.<sup>40</sup>

Certain bacteria have an innate resistance to some antibiotics (*E. coli*, for example, are resistant to vancomycin) that is unique to the biochemistry of the organism but resistance is often acquired from mutation or from the acquisition of external genetic material on plasmids or by other means.<sup>40</sup> Once resistance genes have been acquired, various biochemical mechanisms confer resistance. These may involve the expression of enzymes that hydrolyze or modify an antibiotic so that it becomes ineffective (e.g.  $\beta$ -lactamases that hydrolyses penicillin or enzymes that modify fluoroquinolone so that they no longer inhibit nucleic acid synthesis), changes to the selective permeability of a bacterium so that the intracellular concentration of an antibiotic remains too low for it to function or the adaption of alternative metabolic pathways that compensate for the effects of a drug.<sup>40</sup>

Unlike antibiotics that focus on the treatment of infections by selectively killing the responsible organisms, vaccines illicit a pre-emptive immune response (the production of antibodies) to

antigens that are specific to a pathogen. This may confer either short-term or long-term protection and presents an alternative means of managing infections within a population.

## 4.2 Carbohydrate targeted vaccines

The cell surfaces of many bacterial species are covered by capsular polysaccharides, glycolipids and glycoproteins. A large proportion of the cell surface of gram negative bacteria is covered by a glycolipid known as lipopolysaccharide or endotoxin. Lipopolysaccharide is comprised of a lipid A portion that is embedded in the cell membrane, a core oligosaccharide attached to lipid A and an O-polysaccharide portion that is attached to the core oligosaccharide.<sup>32</sup> Bacterial O-polysaccharides may be comprised of chains of a single type of monosaccharide or of repeating units of several monosaccharides.<sup>41</sup> Structural variation between the O-polysaccharides expressed by different bacteria of the same species is the basis for the classification of bacteria into serogroups.<sup>32</sup> Bacterial lipopolysaccharides are antigens and so elicit the production of antibodies from lymphocytes when suitable hosts (such as humans and other mammals) are exposed to them.

So called “T lymphocyte dependent” antigens require the presence of T lymphocyte cells to function, result in antibodies that have a strong binding affinity for the antigens with which they react and confer long lasting protection to the host. “T lymphocyte independent” antigens, by contrast, do not require the presence of T lymphocytes to function, do not produce antibodies with strong binding affinity for antigens and confer short term protection to the host. T lymphocyte independent antigens can be further categorized as type 1 or type 2 antigens. Type 1, T lymphocyte independent antigens induce the formation of B lymphocyte cells that recognise bacterial cell surface antigens in both adults and newborns.<sup>42</sup> Bacterial carbohydrate antigens thus are typically classed as T lymphocyte independent, type 1 antigens.

Glycoconjugate vaccines are prepared by attaching bacterial polysaccharides fragments to suitable proteins and have been shown to elicit an enhanced immune response compared to bacterial polysaccharides alone.<sup>20</sup> These agents have been shown to confer lasting protection against infection by bacterial species including *Streptococcus pneumoniae*, *Neisseria meningitis* and *Shigella dysenteriae*, among others. The effectiveness of such glycoconjugates prepared from the O-polysaccharide repeating unit of *Shigella dysenteriae* and human serum albumin have been shown to be dependent on oligosaccharide chain length and on the level of protein loading.<sup>8</sup>

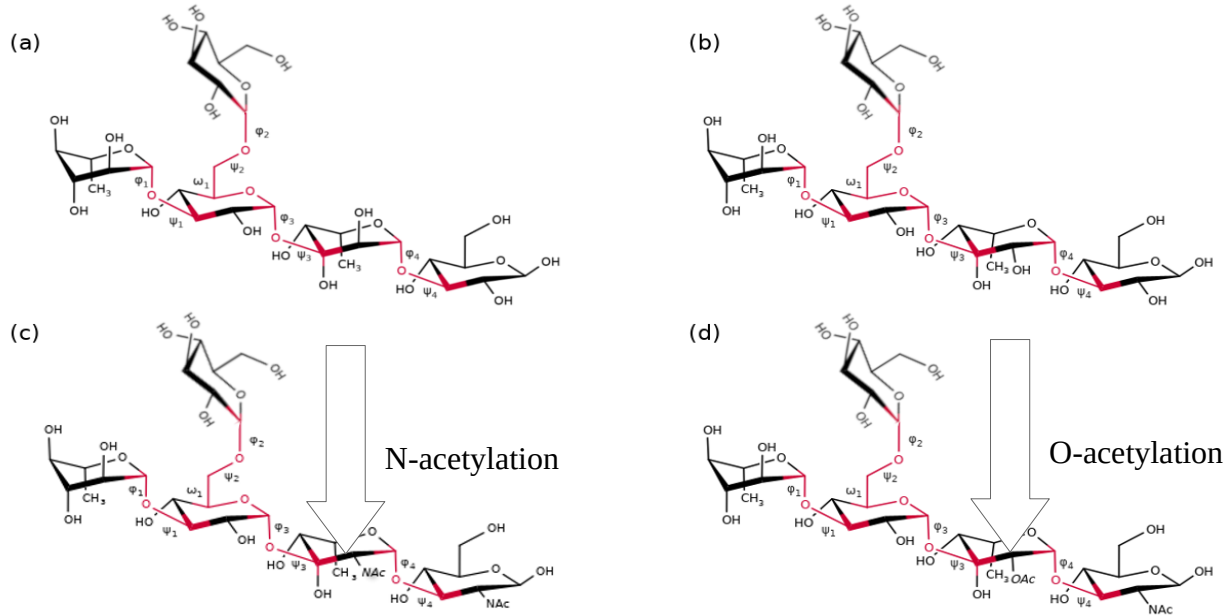
## Chapter 5: *Escherichia coli* serogroups O25 and O25b

*E. coli* is a Gram-negative bacterial species that includes organisms that range from benign inhabitants of the gastrointestinal tract to pathogens that cause life threatening infections.<sup>14</sup> Currently, the species is comprised of bacteria that express more than 70 known capsular polysaccharide antigens, 170 O-antigens (serogroups) and 60 H-antigens.<sup>43</sup> *E. coli* serogroups are commonly linked to particular diseases and are categorized into one or more related pathotypes.<sup>16</sup> Multi locus Sequence Typing (MLST) provides a further means of classifying the panorama of *E. coli* (and other bacteria) based on the sequences of selected portions of the bacterial genome. A sequence type (ST) or “lineage” is assigned by comparing empirically identified genetic information to that stored in an MLST database.<sup>44</sup>

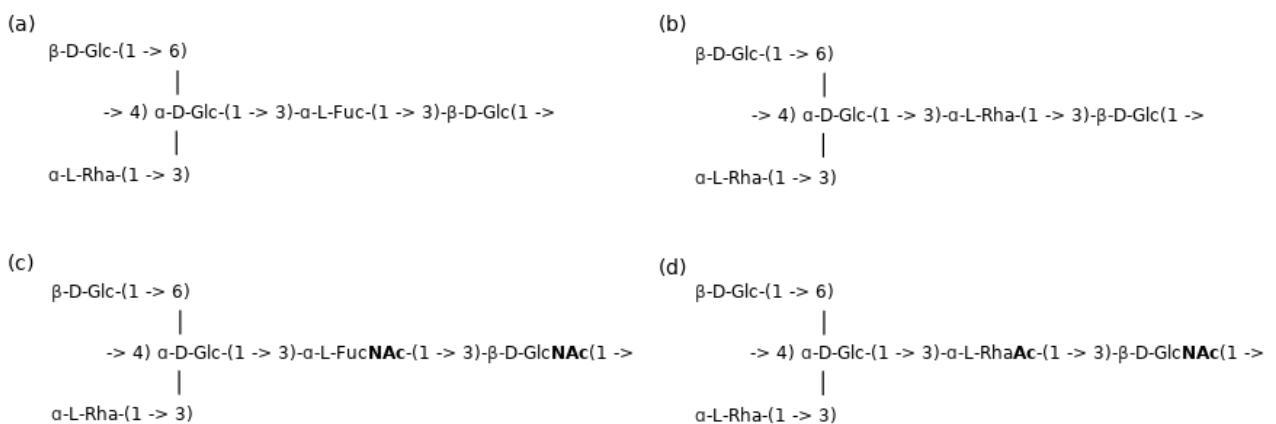
Bacteria from the O25 serogroup, characterized in 1983, are typically enteropathogenic or enterotoxigenic organisms associated with acute intestinal infections that vary in duration and severity.<sup>16</sup> Organisms from the classical O25 serogroup are not particularly remarkable but, in 2008, a previously unknown strain type, ST131, was identified on three continents by MLST of *E. coli* that express a  $\beta$ -lactamase known as CTX-M-15.<sup>45</sup> Subsequent research has confirmed that ST131 is present worldwide and includes organisms that express a broad range of virulence and antibiotic resistance genes.<sup>19</sup> ST131 bacteria make up a large portion of *E. coli* isolates that exhibit resistance to fluoroquinolone and  $\beta$ -lactam antibiotics.<sup>19</sup> Reports of ST131 *E. coli* that are also resistant to carbapenems are concerning as these antibiotics are used as a last resort to treat resistant infections.<sup>46</sup> Most ST131 *E. coli* are of serotype O25b:H4 and thus express the O25b carbohydrate antigen which is related to, but is distinct from, the classical O25 variety that we refer to as O25a for clarity here. ST131:O25b bacteria are often associated with urinary tract and bloodstream infections and are classed into the extraintestinal pathogenic (ExPEC) pathotype.<sup>17</sup> Cases of ST131 serotype O16:H5<sup>47</sup> *E. coli* are also of interest but are less common - the O16 antigen will not be investigated further here.

The high occurrence of infections and antibiotic resistance of ST131-O25b:H4 *E. coli* has resulted in clinical research interest and the lineage has been well characterized.<sup>9,17–19,46–48</sup> Szijarto *et. al.*<sup>9</sup> has determined the molecular structure of the RU of the O25b antigen using empirical methods and contrasted it with that of the previously known O25a variety characterized by Kenne *et. al.*<sup>49</sup>

Figure 7 presents the structures of the O25a and O25b RUs in (c) and (d) respectively. Two additional oligosaccharides that we refer to as O25a' and O25b', and that correspond to the RUs of O25a and O25b without O- or N-acetyl groups, are presented in Figure 7 (a) and (b) respectively. Figure 8 presents the structures of O25a', O25b', O25a and O25b using CASPER notation.



**Figure 7:** Molecular structures of the repeating units of *E. coli* serogroup O25 polysaccharides with and without O-acetyl and N-acetyl groups (a) O25a' (b) O25b' (c) O25a (d) O25b. Glycosidic bonds of interest are highlighted.



**Figure 8:** The monosaccharide sequences of *E. coli* serogroup O25 polysaccharides with and without O-acetyl and N-acetyl groups using CASPER notation (a) O25a' (b) O25b' (c) O25a (d) O25b. O-acetyl and N-acetyl groups are highlighted.

As described in Section 2.5, the orientation of pyranose rings about the glycosidic bonds of an oligo or polysaccharide is the primary determinant of its conformation. The glycosidic bonds of O25a', O25b', O25a and O25b are presented in Table 1 as a series of representative disaccharides that will be referred to throughout the rest of the text.

**Table 1:** Representative disaccharides for the glycosidic bonds of the O25a', O25b', O25a and O25b oligosaccharides and abbreviations used in the rest of the text.

Repeating unit (s)	CASPER Notation	Abbreviation
O25a', O25a O25b', O25b	$\alpha$ -L-Rhap-(1->3)- $\alpha$ -D-Glcp	aR13aG
O25a', O25a O25b', O25b	$\beta$ -D-Glcp-(1->6)- $\alpha$ -D-Glcp	bG16aG
O25a'	$\alpha$ -D-Glcp-(1->3)- $\alpha$ -L-Fucp	aG13aF
O25a'	$\alpha$ -L-Fucp-(1->3)- $\beta$ -D-Glcp	aF13bG
O25b'	$\alpha$ -D-Glcp-(1->3)- $\alpha$ -L-Rhap	aG13aR
O25b'	$\alpha$ -L-Rhap-(1->3)- $\beta$ -D-Glcp	aR13bG
O25a	$\alpha$ -D-Glcp-(1->3)- $\alpha$ -L-FucpNAc	aG13aFNAc
O25a	$\alpha$ -L-FucpNAc-(1->3)- $\beta$ -D-GlcpNAc	aFNAc13bGNAc
O25b	$\alpha$ -D-Glcp-(1->3)- $\alpha$ -L-RhapAc	aG13aRAc
O25b	$\alpha$ -L-RhapAc-(1->3)- $\beta$ -D-GlcpNAc	aRAc13bGNAc
O25a' O25b'	$\beta$ -D-Glcp-(1->4)- $\alpha$ -D-Glcp	bG14aG
O25a O25b	$\beta$ -D-GlcpNAc-(1->4)- $\alpha$ -D-Glcp	bGNAc14aG

## 5.1 Existing structural and conformational studies of *E. coli* O polysaccharides

*E. coli* is a well known species that has been extensively studied<sup>14</sup> but the number of serogroups and the emergence of variants (such as O25b) means that the conformations of all *E. coli* O polysaccharides have not been exhaustively characterized.

A comprehensive review of *E. coli* O polysaccharide structures was presented by Roland Stenutz *et al.*<sup>16</sup> in 2006 and later converted into the ECODAB database<sup>50</sup> which serves as an excellent reference for this information. ECODAB includes an entry containing the molecular structure of the classical O25a antigen (and other information) but does not yet include an entry for the O25b

variety. A search of the database for the aG13aFNAc and aFNAc13bGNAc linkages present in O25a returns results for 7 and 3 known *E. coli* O polysaccharides respectively. A search for the aG13aRAc and aRAc13bG linkages of O25b returns only a single matching result for each linkage: O135 and O16 respectively. Extending this search to include all C1 to C3 linkages that involve C2 O-acetylated rhamnose monosaccharides yields only 1 additional result: O132. This suggests that the aG13aRAc and aRAc13bG linkages of O25b are somewhat unusual among the known *E. coli* O polysaccharides.

GLYCO3D is “a family of databases covering the three-dimensional features of monosaccharides, disaccharides, oligosaccharides, polysaccharides, glycosyltransferases, lectins, mono-clonal antibodies against carbohydrates, and glycosaminoglycan-binding proteins.”<sup>51</sup> In addition to structural information, GLYCO3D includes low energy dihedral values and free energy surfaces for some glycosidic bonds. A review of the GLYCO3D-Polysac database returns entries for the O polysaccharides of *E. coli* 1303, O5ab, O5ac and O65 but does not include entries for either O25a or O25b. An exhaustive search of the GLYCO3D-Disaccharides database for each of the bonds listed in Table 1 only yields results for the aF13bG linkage of O25a.

The Glycosciences.de portal “provides databases and tools to support glycobiology and glycomics research. Its main focus is on 3D structures, including 3D structure models as well as references to PDB entries that feature carbohydrates.”<sup>52</sup> At the time of writing, the GlycoMaps database contains 2585 conformational maps of glycosidic linkages from various oligosaccharides and polysaccharides. A search for the *E. coli* disaccharide fragments presented in Table 1 yields results for aR13aG, aG13aF, aF13bG, aG13aR and bG14aG. An entry for aR13aG (rather than for aR13bG) is also available. These maps were produced from data gleaned from 10 ns molecular dynamics simulations using the MM3 force field in 1996 and present free energy as a function of the dihedral angles,  $\phi$  and  $\psi$ , as we have defined them in Section 2.5. A selection of the surfaces available from Glycosciences.de portal are presented alongside our analogous plots from 100 ns CHARMM36 and GLYCAM06 molecular dynamics simulations in Appendix A for comparison.

Structural studies of *E. coli* O polysaccharides that make use of chemical analysis coupled with Nuclear Magnetic Resonance (NMR) spectroscopy are well documented in the literature and have been carried out for several decades.<sup>9,49,53–60</sup> Stand alone molecular dynamics simulation studies of

*E. coli* O polysaccharides and studies that combine NMR spectroscopy and molecular dynamics simulations are not as common but some have been carried out: For example, Rosen *et. al.* <sup>61</sup> applied the MM3 force field to propose a conformational rationale for antibody cross-reactivity between the *E. coli* O159 and the *Shigella dysenteriae* type 4 O polysaccharides. Steric hindrance between adjacent monosaccharide was noted as an important factor in restricting conformational flexibility in this study. Gosh *et. al.* <sup>62</sup> applied molecular dynamics using the OPLS\_2005 force field alongside NMR spectroscopy during the development of a synthetic method for the *E. coli* O175 repeating unit. Wu *et. al.* <sup>63</sup> used CHARMM36 molecular dynamics simulations and NMR to model the *E. coli* O6 lipopolysaccharide and Blasco *et. al.* <sup>64</sup> applied CHARMM36 to model the *E. coli* O91 polysaccharide. The recently released, impressive, CHARMM-GUI LPS Builder <sup>65</sup> provides an array of data for *E. coli* O antigens but does not yet include structural data for O25b. Remarkably, focused molecular dynamics studies of the conformations of the *E. coli* O25a and O25b polysaccharides are not readily available in the literature.



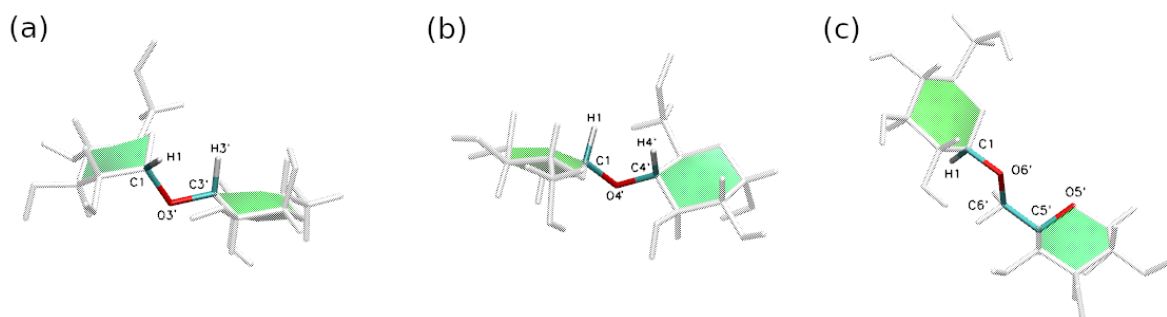
## Chapter 6: Methodology

Our analysis of the O25a and O25b polysaccharide conformations began with the creation of static disaccharide models for each of their constituent glycosidic bonds. Metadynamics<sup>66</sup> simulation of each disaccharide was conducted in vacuum and the resulting Potential of Mean Force (PMF) surfaces were used to prepare 3 RU static oligosaccharide models in aqueous solution. Finally, the 3 RU static models were subjected to unbiased molecular dynamics simulation and data from the resulting trajectories was analysed. In addition to the O25a and O25b oligosaccharide models, analogous 3 RU oligosaccharide extensions of O25a and O25b without acetyl or N-acetyl groups were prepared. These unacetylated models are denoted O25a' and O25b' in the rest of the text for ease of reference and are used to compare the CHARMM36 and GLYCAM06 force fields.

### 6.1 Calculation of preferred disaccharide glycosidic bond dihedrals

Static models of all disaccharides in vacuum were constructed using CarbBuilder v2.0<sup>1</sup>, the psfgen structure building tool and CHARMM36 v1.126 force field parameters<sup>24–28</sup>. The online GLYCAM-Web<sup>37</sup> application and GLYCAM06j force field parameters<sup>29</sup> were used to construct static models of all disaccharides without O- or N-acetyl groups in vacuum (N-acetyl-L-rhamnosamine and N-acetyl-L-fucosamine monosaccharides are not available via GLYCAM-Web at the time of writing).

Biased MD simulations of each disaccharide in vacuum were performed using NAMD v2.13<sup>67</sup> and Metadynamics<sup>66</sup>. Glycosidic bond dihedrals were defined as  $\varphi = \{H_1 - C_1 - O_x' - C_x'\}$ ,  $\psi = \{C_1 - O_x' - C_x' - H_x'\}$  and  $\omega = \{O_6' - C_6' - C_5' - O_5'\}$  and set as collective variables (Figure 9). Each simulation was run for 100ns with a hill height and width of 0.05 kcal/mol and 2.5 degrees.



**Figure 9:** Disaccharide models with the atoms defining  $\varphi$ ,  $\psi$  and  $\omega$  dihedral angles for (a) C1 to C3, (b) C1 to C4 and (c) C1 to C6 glycosidic bonds respectively.

## 6.2 Construction of aqueous phase 3 RU oligosaccharides models

### 6.2.1 CHARMM models

Four static 3 RU oligosaccharide models of O25a and O25b (both with and without O- and N-acetyl groups) in vacuum were constructed using CarbBuilder v2.0, the psfgen structure building tool and CHARMM36 v1.126 force field parameters. CarbBuilder was configured to use preferred glycosidic bond dihedrals based on the PMF contour plots described in Sections 6.1 and 6.4. A solvent box of TIP3P water molecules with dimensions of 70 x 70 x 70 Å was added to each model using the VMD v1.9.3<sup>68</sup> solvate and auto-ionize functions.

### 6.2.2 GLYCAM models

Two static 3 RU oligosaccharide models of O25a' and O25b' (without O-acetyl or N-acetyl groups) in vacuum were constructed using the online GLYCAM-Web application and GLYCAM06j force field parameters. GLYCAM-Web was configured to use preferred glycosidic bond dihedrals based on the PMF contour plots described in sections 6.1 and 6.4. A neutralized solvent box of TIP3P water molecules with dimensions of 70 x 70 x 70 Å was added to each model using the AmberTools<sup>69</sup> tleap application. GLYCAM-Web includes a solvate function but it does not appear to allow precise specification of water box dimensions (only provides a “padding function”) and did not produce satisfactory solution phase models. In addition, oligosaccharide structure files (.parm7 files) produced by GLYCAM-Web were not compatible with NAMD and had to be re-built from the respective co-ordinate (.pdb files) with tleap prior to molecular dynamics simulation.

## 6.3 Molecular dynamics simulations

Computations were performed using the facilities provided by the University of Cape Town's ICTS High Performance Computing team. All molecular dynamics simulations were performed using v2.13 of the NAMD molecular dynamics application. Force field simulations used either the CHARMM36 v1.126 force field parameter set or the GLYCAM06j force field parameter set as appropriate for the models described in Section 6.2. NAMD was configured as follows:

Metadynamics simulations of representative disaccharide fragments in vacuum were performed at 310K and were preceded by 10 000 steps of minimization in each case. A heating protocol consisting of 5K increments from 10K up to 300K with 500 steps of minimization at each increment was applied to the solution oligosaccharide extension simulations. Equations of motion were integrated using a step size of 1 femtosecond in both the vacuum and solution simulations. Solution

simulations were performed under constant temperature and pressure, used periodic boundary conditions (cubic, 70 Å cells), applied a dielectric constant of 1 and used Particle Mesh Ewald (PME) summation for electrostatic interactions. A switching function was applied between 12.0 and 15.0 Å to smooth the truncation of non-bonded interactions in the solution phase simulations. Vacuum disaccharide simulations were run for 100 ns each. Solution phase simulations of the oligosaccharide extensions were run for 200ns each. An example NAMD configuration for a vacuum disaccharide simulation (showing all parameters as well as the selection of collective variables for Metadynamics) is provided in Appendix B and C. An example NAMD configuration for a solution phase oligosaccharide simulation (showing all parameters) is provided in Appendix E.

## 6.4 Data analysis

### 6.4.1 *Biased (Metadynamics) simulations in vacuum*

PMF files from Metadynamics simulations of disaccharides in vacuum were graphed and contoured in batches using custom bash and Python scripts in conjunction with Gnuplot <sup>70</sup> v5.2 patch level 6 and ImageMagick <sup>71</sup> v6.9.7-4. The custom Python script searches for the lowest energy dihedral values from PMF files and makes the data point available to Gnuplot for representation on PMF contours via a bash script. GNU Image Manipulation Program (GIMP) v2.8.22 was used to combine graphs and create custom graphics.

### 6.4.2 *Unbiased solution phase simulations*

Unbiased solution phase NAMD simulations produced large trajectory files that included hundreds of thousands of frames (several hundred gigabytes of data per simulation). Extracting and analyzing time series information for selected variables from these cumbersome files required a well engineered approach. In all cases, we have disregarded the first 100 ns of simulation time for equilibration of the molecules and focus our analysis on the last 100 ns.

#### 6.4.2.1 *Removing solvent data from trajectory files*

Much of the storage capacity consumed by solution phase trajectory files holds information about solvent molecules. This is, of course, important while simulations are in progress but was no longer explicitly required for this study once they were completed. v5.1 of the Catdcd application (included with VMD v1.9.3) was used to extract trajectory information for the oligosaccharide

molecules under study into solvent free trajectory files prior to further analysis. This technique produced reduced trajectory files that were approximately ten times smaller than those that included solvent information. This made further trajectory analysis much less resource intensive.

#### 6.4.2.2 *Extracting time series glycosidic bond dihedrals from trajectories*

Extracting and analyzing time series values for the glycosidic bond dihedral values of each oligosaccharide from trajectory files required automation. A custom TCL library, used in conjunction with VMD, was written by the author over the course of this study to achieve this. The library includes procedures that allow for the selection of multiple dihedral angles of interest from the VMD molecule viewer and outputs their values over a selected range of frames to a labeled file. An equivalent procedure that takes predefined dihedral name and atom index files as arguments is included to facilitate using VMD in text mode. This feature, in conjunction with custom bash scripts, allowed for expedient data extraction using the UCT HPC Cluster (where simulations were run) and avoided the need to download large trajectory files onto a desktop computer.

#### 6.4.2.3 *Probability density and scatter plots of glycosidic bond dihedrals*

Probability density plots for glycosidic bond dihedrals were graphed from the time series data files produced using VMD and the TCL script described in Section 6.4.2. Custom Python code was written to select, read and plot the data as histograms using the Numpy<sup>72</sup> and Matplotlib<sup>73</sup> libraries. As many probability density plots were graphed over the course of this study, a simple GUI was created to simplify the use of this code using the Tkinter library.

The data files described in Section 6.4.2 were used to produce scatter plots of time series glycosidic bond dihedrals. Scatter plots were graphed and overlaid with the PMF contour plots described in Section 6.1 using the Gnuplot application with custom script.

#### 6.4.2.4 *Time series and probability density plots of 3 RU oligosaccharide end-to-end distances*

The distance between the ring oxygen atoms of the first and last  $\alpha$ -D-glycopyranose monosaccharides of each oligosaccharide extension were extracted from simulation trajectory files using a method analogous to that described in Sections 6.4.2.2 and 6.4.2.3 for dihedral angles. A TCL procedure was written to extract distance values for a selected series of frames into files and

Python code was written to plot the data using the Numpy and Matplotlib libraries. The Tkinter GUI, described previously, was used to make it easier to select files and plot the data.

#### 6.4.2.5 *Hydrogen bond analysis*

Hydrogen bond analysis was carried out using the VMD Hydrogen Bond plugin with a cutoff distance setting of 3.5 Å and a cutoff angle setting of 30 degrees (as set in VMD).

#### 6.4.2.6 *Root Mean Square Deviation (RMSD) cluster analysis*

RMSD clustering analysis was performed using the Physbio Clustering Tool (Tcl script / VMD plugin) written Luis Gracia, Cornell University Medical College. 4 clusters were selected for all analysis. Samples were taken every 2.5 ps for the last 100 ns of each simulation. As carbohydrates are fairly flexible, a cluster cutoff value of 5 was applied for analysis of the O25a' and O25b' trimers. A cutoff value of 4.5 was applied to analysis of the O25a and O25b trimers.

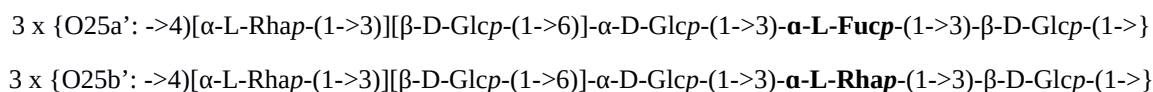
## Chapter 7: Results and discussion

We begin with a review of the level of extension and glycosidic bond dihedral values of 3 RU oligosaccharides extensions of O25a and O25b without O- or N-acetylation (O25a' and O25b'). Data from our comparative, solution phase simulations of trimers of O25a' and O25b' is presented as a case study comparing the CHARMM36 and GLYCAM06 force fields. Thereafter, we compare the conformations of 3 RU oligosaccharide extensions of O25a and O25b (including O- and N-acetyl groups) in Section 7.2. Only CHARMM36 data is presented for the trimers of O25a and O25b - GLYCAM06 simulations of these substituted molecules were not performed.

### 7.1 CHARMM36 and GLYCAM06: Force field comparison

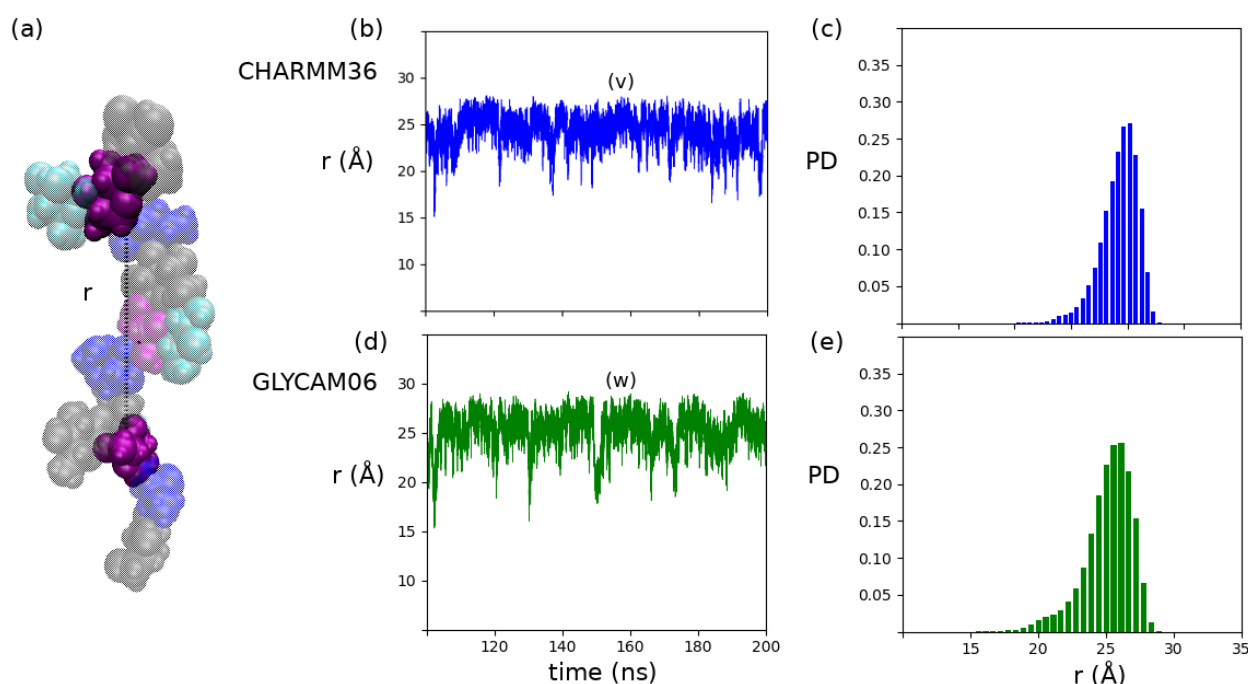
#### 7.1.1 Preferred level of extension of O25a' and O25b' trimers

The distance ( $r$ ) between the ring oxygen atoms of  $\alpha$ -D-glucopyranose monosaccharides nearest to the reducing and non-reducing ends of each oligosaccharide trimer are used as a measure of their level of extension (Figure 10 a). We begin with a comparison of  $r$  values from CHARMM36 and GLYCAM06 simulations of the O25a' and O25b' trimers:



The first 100 ns of each simulation is disregarded for equilibration leaving 100 ns of simulation time for analysis. A time series plot of  $r$  for O25a' calculated from CHARMM36 data (Figure 8 b) shows rapid transitions between 15.1 Å and 28.1 Å with an average of 24.3 Å (standard deviation of 1.6 Å). The GLYCAM06 data (Figure 10 d) suggests a marginally longer O25a' oligosaccharide with transitions between 15.3 Å and 29.2 Å and an average of 25.1 Å (standard deviation of 1.8 Å).

Probability density plots of  $r$  (Figure 10 c and e) using data from both force fields are positively skewed distributions. The CHARMM36 data presents a well differentiated maximum of 0.27 at about 25.2 Å whereas the GLYCAM06 data suggests a slightly longer, slightly more flexible oligosaccharide with a maximum of 0.26 at about 26.1 Å.

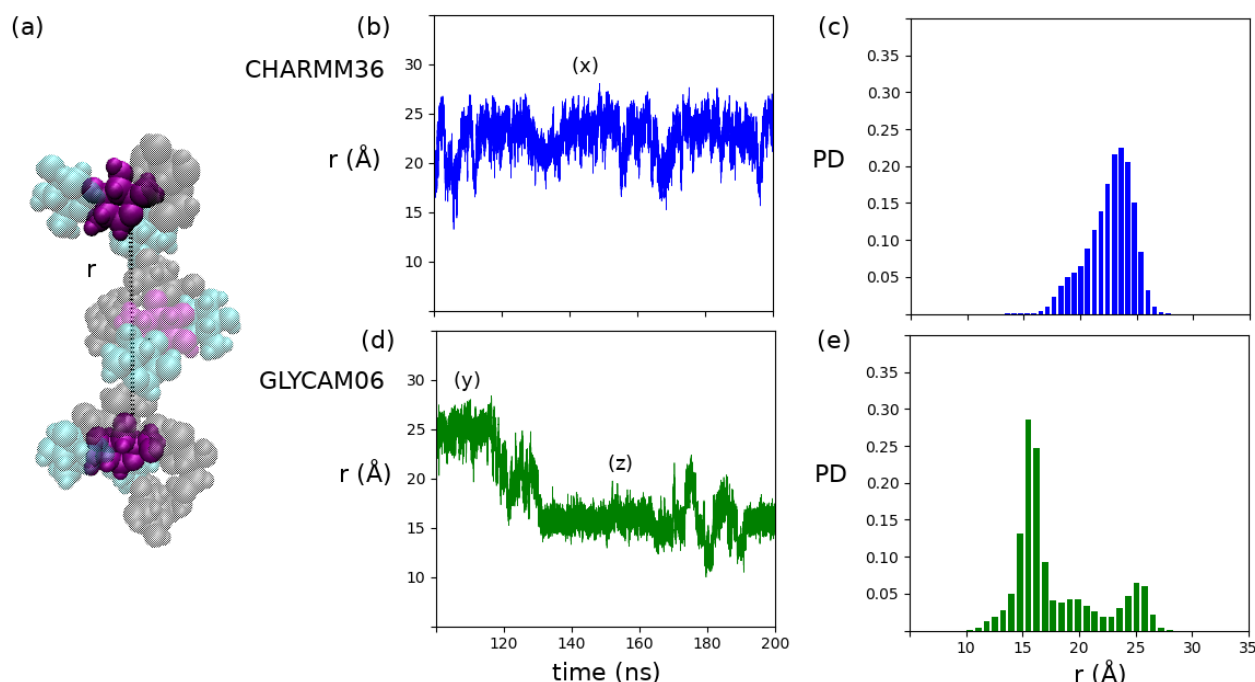


**Figure 10:** (a) Model of the O25a' trimer showing the selected measure of extension,  $r$ .  $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc in grey,  $\alpha$ -L-Rha in cyan, and  $\alpha$ -L-Fuc in blue. (left) Time series end-to-end distance plots of (b) O25a' using CHARMM36 (d) O25a' using GLYCAM06. (v) and (w) represent the positions of conformers illustrated in Figure 10. (right) Probability density plots for the end-to-end distance of (c) O25a' using CHARMM36 and (e) O25a' using GLYCAM06.

Overall, CHARMM36 and GLYCAM06 are in reasonable agreement regarding the level of extension of the O25a' oligosaccharide extension with  $\Delta r = 0.9 \text{ \AA}$  for the most probable conformations. O25b' differs from O25a' in that the central N-acetyl- $\alpha$ -L-fucosamine unit in O25a' is replaced with 2 O-acetyl- $\alpha$ -L-rhamnopyranose in O25b'. Comparison of  $r$  values for O25a' and O25b' reveal that this difference appears to affect the level of extension of the oligosaccharides, as follows.

A time series plot of  $r$  for O25b' calculated from CHARMM36 data (Figure 9 b) shows rapid transitions between  $13.2 \text{ \AA}$  and  $28.1 \text{ \AA}$  with an average of  $22.6 \text{ \AA}$  (standard deviation of  $2.0 \text{ \AA}$ ). A CHARMM36 probability density plot of  $r$  for O25b' (Figure 11 c) is broader and shorter than that for O25a' with a maximum of 0.22 at about  $23.5 \text{ \AA}$ . These results indicate that CHARMM36 predicts a more compressed conformation of the O25b' trimer than it does for the O25a' trimer with  $\Delta r = 1.7 \text{ \AA}$  between the most probable conformations of the two oligosaccharides.

CHARMM36 and GLYCAM06 predict substantially different levels of extension for O25b' in solution. The GLYCAM06 end-to-end distance plot presented in Figure 11 (d) indicates that transitions occur between  $r$  values of 9.97 Å and 28.4 Å with an average of 18.0 Å (standard deviation of 3.8 Å). The GLYCAM06 probability density plot of  $r$  (Figure 11 e) exhibits a very prominent primary maximum of 0.29 at about 15.5 Å. The presence of a much lower secondary maximum of 0.065 at about 25.1 Å is interesting and indicates that the O25b' conformation collapses into a more compact conformation after about 116ns of simulation with GLYCAM06. This transition is not reversible over the remainder of the simulation duration. The difference in the preferred levels of extension of O25b' predicted by CHARMM36 and GLYCAM06 is thus very large with  $\Delta r = 8.0$  Å for the most probable values.



**Figure 11:** (a) Model of the O25b' trimer showing the selected measure of extension,  $r$ . α-D-Glc in purple, β-D-Glc in grey, α-L-Rha in cyan. (left) Time series end-to-end distance plots of (b) O25b' using CHARMM36 (d) O25b' using GLYCAM06. (x), (y) and (z) represent the positions of conformers illustrated in Figure 11. (right) Probability density plots for the end-to-end distance of (c) O25b' using CHARMM36 and (e) O25b' using GLYCAM06.

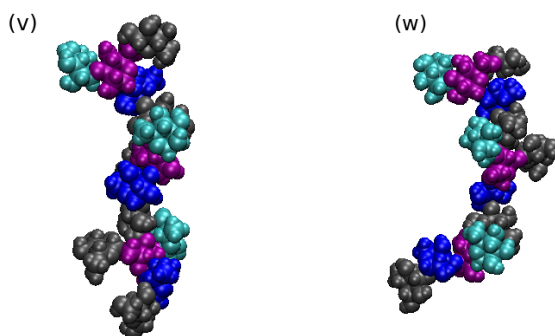
Calculated values of  $r$  for the unacetylated, 3 RU O25a' and O25b' oligosaccharide extensions are summarized in Table 1. We now turn our attention to the predicted conformations of the O25a' and O25b' trimers to rationalize the observed differences in their preferred levels of extension.



**Table 1:** The level of extension of O25a' and O25b' presented as the end-to-end distance between the ring oxygen atoms of their reducing and non-reducing  $\alpha$ -D-glucopyranose units ( $r$ ). Minimum, maximum, average and most probable values calculated from CHARMM36 and GLYCAM06 simulation data are indicated. Standard deviations (for averages) and probability density peak heights (for most probable values) are bracketed.

	CHARMM36				GLYCAM06				Delta
	Min. $r$ (Å)	Max. $r$ (Å)	Avg. $r$ (Å)	Prob. $r$ (Å)	Min. $r$ (Å)	Max. $r$ (Å)	Avg. $r$ (Å)	Prob. $r$ (Å)	$\Delta r$ (Å)
O25a'	15.1	28.1	24.3 (1.6)	25.2 (0.27)	15.3	29.2	25.1 (1.8)	26.1 (0.26)	0.9
O25b'	13.2	28.1	22.6 (2.0)	23.5 (0.22)	9.97	28.4	18.0 (3.8)	15.5 (0.29)	8

The two force fields are largely in agreement regarding the level of extension of the O25a' trimer. Representative conformers that correspond with the most probable  $r$  values calculated from the CHARMM36 and GLYCAM06 simulations are illustrated in Figure 10 v and w respectively (Refer to the points marked v and w on Figure 10). The glycosidic bond dihedrals of the central RU of these two conformers all fall within the respective low energy conformational regions that we present in Section 7.1.2. Additional O25a' conformations that occur over the analyzed simulation time are presented in Appendix F.



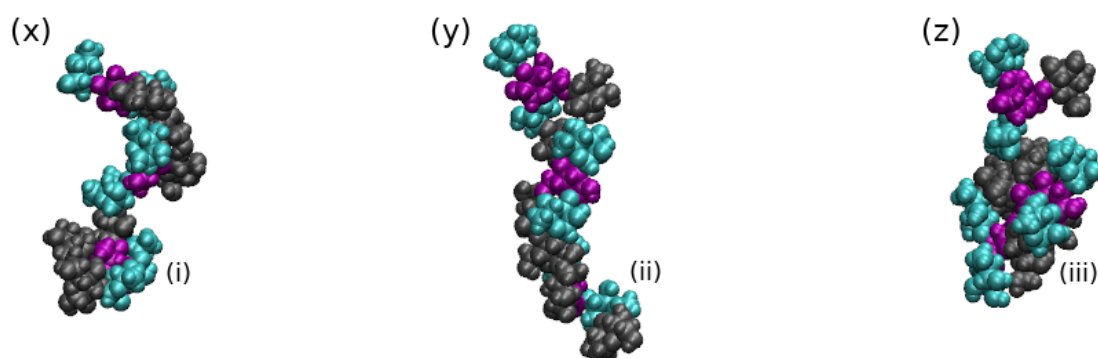
**Figure 12:** Representative conformations of O25a' extracted from the trajectories of (v) the CHARMM36 simulation where  $r = 25.2$  Å at 155.754 ns and (w) the GLYCAM06 simulation where  $r = 26.1$  Å at 154.072 ns. ( $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc in grey,  $\alpha$ -L-Rha in cyan, and  $\alpha$ -L-Fuc in blue)

The GLYCAM06 plot of  $r$  for O25a' (Figure 10 d) presents a notable trough between about 149 ns and 152 ns. Some shorter compressions of the oligosaccharide are evident in the analogous CHARMM36 plot (Figure 10 b) but the molecule tends to resume a more extended conformation

quite quickly. A representative O25a' conformer extracted from the GLYCAM06 simulation trajectory at 150.38 ns (Appendix F, Figure F1 g) shows substantial folding of the RU at the reducing end of the oligosaccharide towards the central RU. This configuration is conducive to the formation of multiple intramolecular hydrogen bonds that may serve to stabilize the collapsed conformation. Although the O25a' oligosaccharide returns to a more extended arrangement at about 153ns, this feature may be related to the pronounced, irreversible transition of O25b' into a collapsed conformation during our GLYCAM06 simulation (Figure 11 d).

GLYCAM06 has a documented tendency to favor more compact oligosaccharide conformations than CHARMM36 that may be due to intra/inter molecular interactions dominating over electrostatic interactions.<sup>22</sup> This presents a possible explanation for the observed trough in the GLYCAM06 plot of O25a' (Figure 10 d): once certain folded conformations form, the molecule takes longer to re-extend with GLYCAM06 than with it does with CHARMM36. On average, this feature did not have a large impact on  $\Delta r$  for O25a' - in fact, GLYCAM06 predicts a slightly more extended conformation of the O25a' trimer than CHARMM36 does overall. The occupation of compact conformers is considered to be low for the 3 RU O25a' oligosaccharide.

Unlike those for O25a', our comparative simulations of O25b' indicate that GLYCAM06 predicts a much more compressed oligosaccharide than CHARMM36 overall. O25b' conformers that correspond to the points labeled (x), (y) and (z) marked on the time series plots in Figure 9 b and d are presented in Figure 11.



**Figure 13:** Representative conformations of the O25b' trimer extracted from the trajectories of: (x) the CHARMM36 simulation where  $r = 23.6 \text{ \AA}$  at 145.058 ns, (y) the GLYCAM06 simulation where  $r = 26.08 \text{ \AA}$  at 114.534 ns and (z) the GLYCAM06 simulation where  $r = 15.83 \text{ \AA}$  at 148.815 ns.

Note the folding of the oligosaccharides near to their reducing ends at (i), (ii) and (iii) respectively. ( $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc in grey and  $\alpha$ -L-Rha in cyan)

Kang *et. al.*<sup>74</sup> observed “extreme, hairpin-like conformations” for oligosaccharide extensions of the O-antigen of *Shigella flexneri* Serotype Y in a comparative study using the CHARMM36 and GLYCAM06 force fields. The formation of these hairpin-like conformers led to metastable compact states stabilized by intramolecular bonds in some GLYCAM06 simulations. These metastable states were not observed in analogous CHARMM36 simulations.

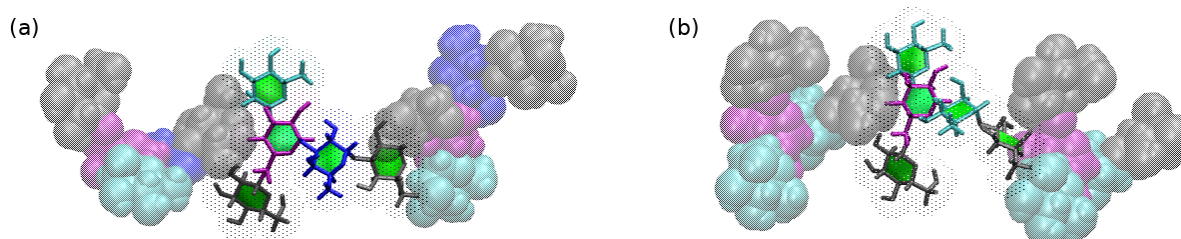
Our comparative simulations of the 3 RU O25b’ oligosaccharide extension align well with the analysis presented by Kang *et. al.*<sup>74</sup>. In our GLYCAM06 simulation, the trimer begins in an extended conformation at about 114.5 ns (Figure 13 y). The non-reducing RU folds back towards the center of the molecule forming a hairpin-like arrangement around a  $\beta$ -D-glucopyranose side chain and reducing the value of  $r$ . Finally, dominating attractive intramolecular forces in GLYCAM06 result in the collapse of the molecule into a metastable state represented in Figure 13 z. This compact conformation appears to vary somewhat in extension at around 180 ns but dominates for the majority of the assessed simulation time. It is worth noting that the collapse of the trimer only began after about 115 ns of simulation and that a much more extended conformation was favored for the first 100 ns (Refer to Appendix G, Figure G2). Shorter simulations may have missed this behavior of GLYCAM06 entirely - the results of which on the outcome of the simulation are unclear. This serves to emphasize the important general finding that long simulation times are needed to expose the conformational behavior of flexible oligosaccharides in solution as noted by Kuttel *et. al.*<sup>22</sup> Our analysis of data gleaned from CHARMM36 simulation of O25b’ does not preclude the formation of hairpin-like arrangements but, if they do form, they do not lead to the collapse of the molecule into a metastable, compact conformation. A representative O25b’ conformer from our CHARMM36 simulation is presented in Figure 12 v.

Several key points can be taken from our comparative investigations of the level of extension of O25a’ and O25b’ trimers: First, despite some differences, both CHARMM36 and GLYCAM06 predict that an extended conformation of O25a’ dominates in solution. Second, CHARMM36 and GLYCAM06 predict markedly different end-to-end distances for the O25b’ trimer. GLYCAM06 predicts that a collapsed conformation predominates whereas CHARMM36 favors a much more

extended arrangement. Finally, the  $\beta$ -D-glucopyranose side chain of the central RU of the O25b' trimer appears very constrained in the collapsed conformation predicted by GLYCAM06.

### 7.1.2 Preferred glycosidic bond dihedrals for O25a' and O25b'

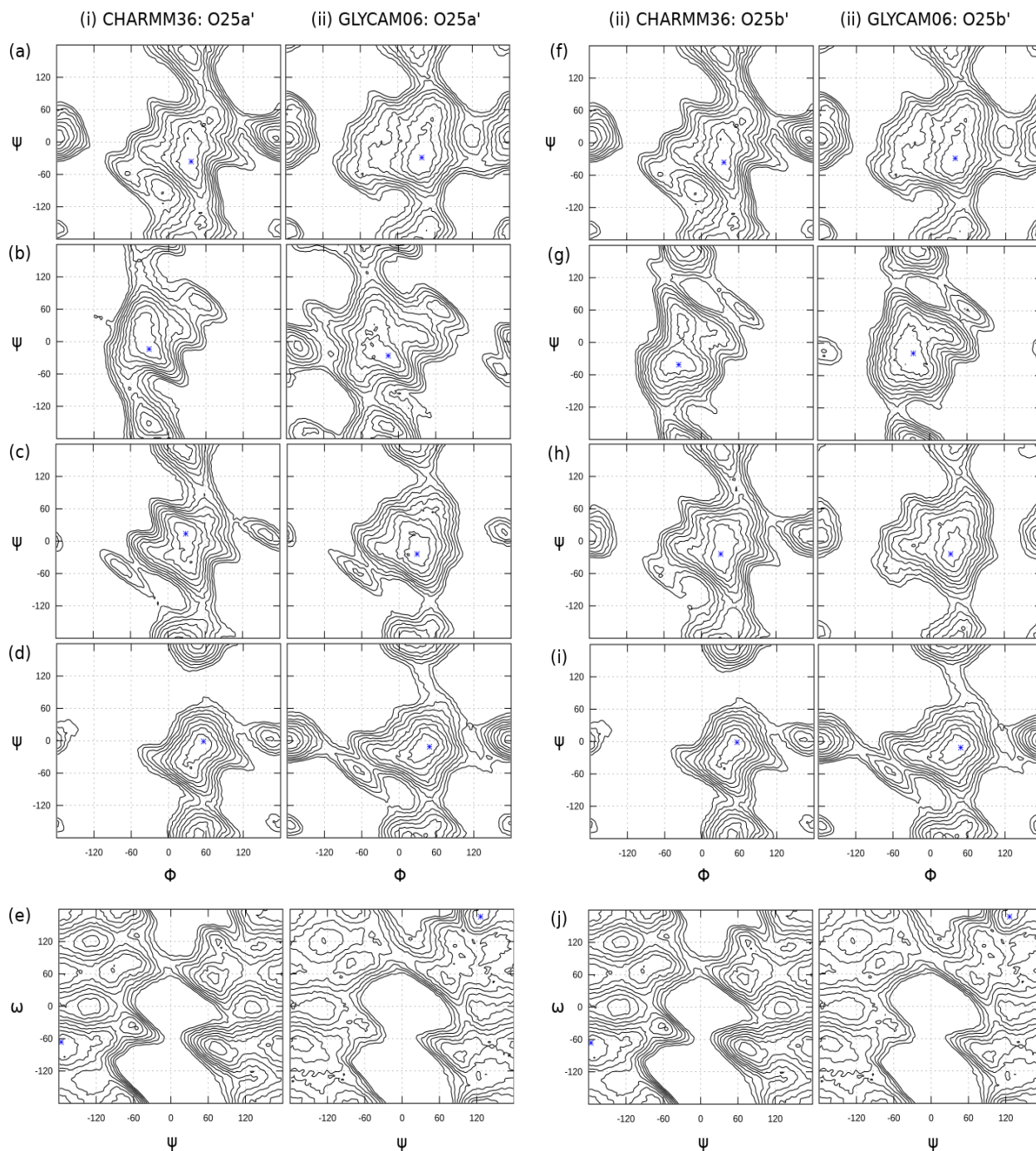
We now turn to an analysis of the dihedral angles of the constituent glycosidic bonds of the O25a' and O25b' trimers to further rationalise the differences described in Section 7.1.1. The glycosidic bonds of the central RU of each oligosaccharide, illustrated in Figure 14, are considered to be a better representation of the conformation of a long polysaccharide than those at the ends. Future references to the O25a' and O25b' RUs refer to this central pentasaccharide.



**Figure 14:** Minimized oligosaccharide models highlighting the glycosidic bonds of the central RUs of the (a) O25a' and (b) O25b' trimers. Note the absence of O- and N-acetyl groups. ( $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc in grey,  $\alpha$ -L-Rha in cyan, and  $\alpha$ -L-Fuc in blue)

The conformations of polysaccharides are generally considered to be determined by their glycosidic bond dihedral angles rather than by inter-residue interactions. As such, low energy oligosaccharide conformations can be estimated by considering the preferred orientations of their glycosidic linkages in isolation.<sup>22</sup> The use of relatively low energy starting structures for molecular dynamics simulations of oligosaccharides is important as high energy starting structures will result in invalid populations of conformers.<sup>23</sup>

Potential of Mean Force (PMF) energy surfaces of vacuum, Metadynamics simulation data for disaccharides that represent the constituent glycosidic bonds of O25a' and O25b' are presented in Figure 17. The low energy regions of these PMF surfaces were used to inform our selection of dihedral angles for the starting structures of O25a' and O25b' oligosaccharide extensions. Data from both CHARMM36 and GLYCAM06 Metadynamics simulations is presented for each representative disaccharide here. Additional analysis of these PMF surfaces and comparison with MM3 force field data from the literature<sup>52</sup>, is presented in Appendix A.



**Figure 15:** PMF surfaces plotted from (i) CHARMM36 and (ii) GLYCAM06 vacuum molecular dynamics simulation data for representative disaccharides of the RUs of O25a' and O25b'. Data for the following bonds are presented: for O25a' (a) aR13aG (b) aG13aF (c) aF13bG (d) bG14aG (e) bG16aG, and for O25b': (f) aR13G (g) aG13aR (h) aR13bG (i) bG14aG (j) bG16aG. Contour lines are plotted using increments of 1 kcal/mol up to a maximum cutoff value of 9 kcal/mol. Points that represent global minimum energy values are presented as blue stars.

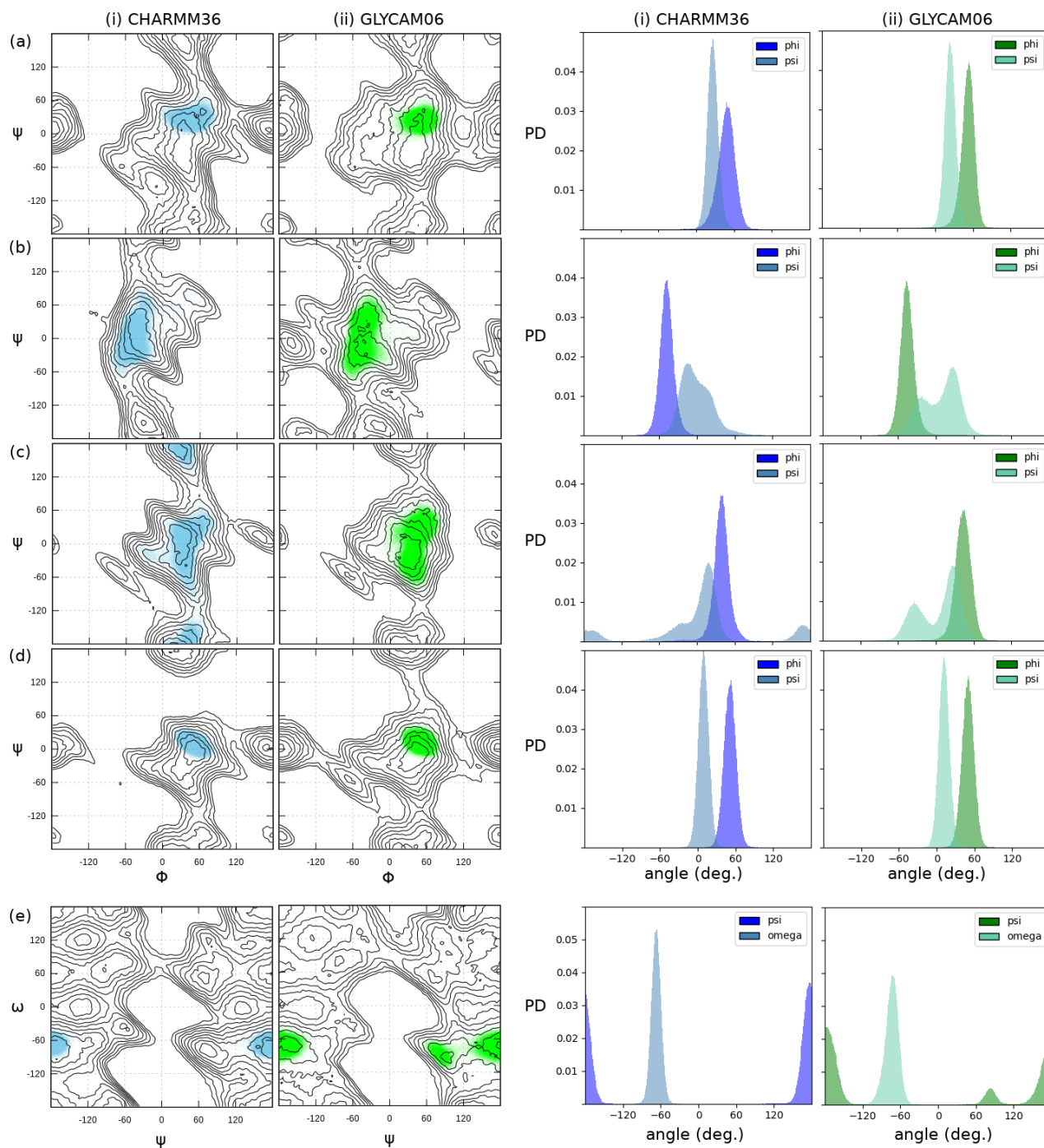


Dihedral angles for the starting structures of the O25a' and O25b' trimers were set as close to the PMF surface primary energy minima as possible. Two deviations from this approach should be noted: First, the region around the non-reducing end of the  $\alpha$ -D-glucopyranose monosaccharide of the central RUs is quite constrained (Figure 14). The dihedral angles of the aR13aG bond (Figure 13 a and f) that involves this unit were initially set to the local syn-anti PMF minimum for both O25a' and O25b'. Second, the dihedral angles of the aG16aG bond of O25a' for the GLYCAM06 simulation started at  $\phi = 42$ ,  $\Psi = 43$  and  $\omega = 171$  which are different to the PMF minima presented for this bond. This may have been due to the GLYCAM-Web minimization feature. Review of the data presented in Figures 16 and 17 indicates that these bonds reached low energy conformations and neither of these deviations appear to have had an effect on the outcome of the simulations.

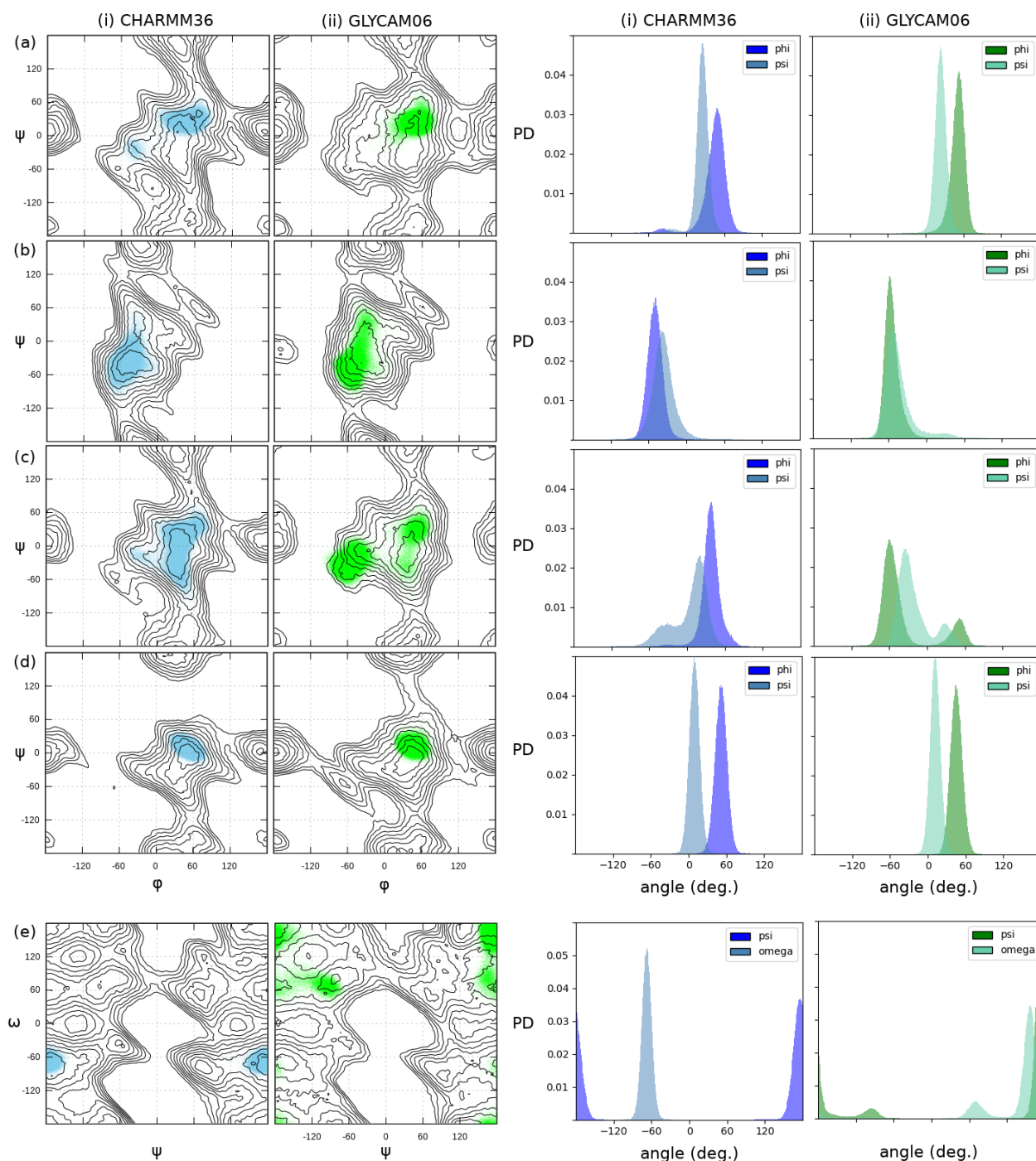
PMF global energy minima for CHARMM36 and GLYCAM06 are in agreement for the aR13aG and bG14aG bonds common to both RUs. There are some differences between the CHARMM36 and GLYCAM06 global PMF energy minima for the aG13aF, aF13bG and aG13aR bonds (Figure 13 b, c and g). Our CHARMM36 values for the aG13aR dihedrals ( $\phi = -36^\circ$ ,  $\Psi = -41^\circ$ ) are in reasonable agreement with those reported in a CHARMM-GLYCAM force field comparison study of *Streptococcus pneumoniae* antigens that used longer Metadynamics simulation times ( $\phi = -33.75^\circ$ ,  $\Psi = -38.75^\circ$ )<sup>2</sup>. Our GLYCAM values for this bond ( $\phi = -26^\circ$ ,  $\Psi = -19^\circ$ ) however, are not in agreement with those reported in this study ( $\phi = -13.75^\circ$ ,  $\Psi = -53.75^\circ$ )<sup>2</sup>. Longer Metadynamics simulations may be useful in verifying these differences.

Overall, the differences between the low energy regions of the CHARMM36 and GLYCAM06 PMF surfaces do not appear to be large enough to explain the large difference in the preferred levels of extension predicted for the O25b' trimer by the two force fields. Inter- or intramolecular interactions appear to play an important role here.

We now move to a review of unbiased, aqueous phase simulation data for the O25a' and O25b' RUs to investigate the preferred dihedral angles occupied by their glycosidic bonds in solution. We have disregarded the first 100 ns of simulation time for equilibration leaving 100 ns for analysis. Scatter and probability density plots of CHARMM36 and GLYCAM06 data for the central RUs of each oligosaccharide are presented in Figures 16 and 17 respectively. The scatter plots have been overlaid onto the disaccharide PMF surfaces presented in Figure 15.



**Figure 16:** (left) Scatter plots and (right) probability density plots representing the preferred dihedral angle value populations of the central RU of the O25a' trimer. (i) CHARMM36 and (ii) GLYCAM06 data is presented for (a) aR13aG (b) aG13aF (c) aF13bG (d) bG14aG and (e) bG16aG. Note: The restriction of rotation around the (a) aR13aG, (d) bG14aG and (e) bG16aG glycosidic bonds; (c) CHARMM36 indicates syn-anti conformer populations that are absent in GLYCAM06. Contour lines are increments of 1kcal/mol up to a cutoff of 9 kcal/mol.



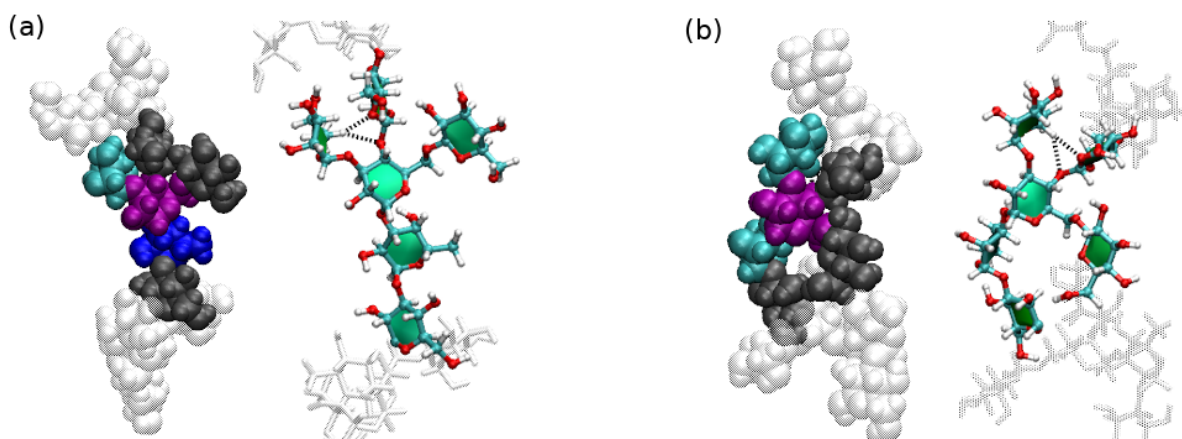
**Figure 17:** (left) Scatter plots and (right) probability plots for dihedral angle populations of the central RUs of the O25b' trimer. (i) CHARMM36 and (ii) GLYCAM06 data for (a) aR13aG (b) aG13aR (c) aR13bG (d) bG14aG and (e) bG16aG. Note: (a) There is a small population of conformers at  $\phi \sim -40^\circ$  for CHARMM36; (b) GLYCAM06 and CHARMM36 populations are shifted slightly; (c) A population of conformers at  $\phi \sim -60^\circ$  is present for aR13bG in GLYCAM06 (e) Preferred conformers for the bG16aG bond are different for CHARMM36 and GLYCAM06. Contour lines are increments of 1kcal/mol up to a cutoff of 9 kcal/mol.



**Table 2:** Preferred dihedral angle values calculated from probability density plots of the constituent glycosidic bonds of the O25a' and O25b' RUs. Probability density peak heights are indicated in brackets next to their respective dihedral angle values.

RU	Glycosidic Bond	CHARMM36 dihedral angles			GLYCAM06 dihedral angles		
		$\phi / ^\circ$	$\Psi / ^\circ$	$\omega / ^\circ$	$\phi / ^\circ$	$\Psi / ^\circ$	$\omega / ^\circ$
O25a'	aR13aG	48.4 (0.032)	26.4 (0.048)	-	53.0 (0.042)	23.1 (0.048)	-
O25a'	aG13aF	-47.1 (0.039)	-13.5 (0.018)	-	-47.3 (0.040)	25.3 (0.018) -22.5 (0.0010)	-
O25a'	aF13bG	39.1 (0.038)	19.5 (0.020) 169.5 (0.0042)	-	42.2 (0.034)	27.3 (0.019) -34.4 (0.0010)	-
O25a'	bG14aG	51.0 (0.043)	8.9 (0.050)	-	50.4 (0.043)	10.6 (0.049)	-
O25a'	bG16aG	-	176.5 (0.037)	-66.3 (0.054)	-	179.6 (0.024) 83.5 (0.0049)	-71.8 (0.040)
O25b'	aR13aG	48.1 (0.032) -36.5 (0.0011)	26.1 (0.048) -21.0 (0.001)	-	53.1 (0.041)	23.3 (0.047)	-
O25b'	aG13aR	-48.2 (0.036)	-40.0 (0.028)	-	-58.5 (0.041)	-52.3 (0.028) 24.9 (0.0013)	-
O25b'	aR13bG	37.5 (0.037)	21.1 (0.023) -30.5 (0.006)	-	-59.1 (0.028) 51.2 (0.0070)	-36.0 (0.025) 28.2 (0.0057)	-
O25b'	bG14aG	51.4 (0.043)	8.0 (0.05)	-	44.8 (0.043)	11.7 (0.050)	-
O25b'	bG16aG	-	176.5 (0.037)	-67.5 (0.054)	-	171.5 (0.047) -96.5 (0.003)	157.5 (0.034) 67.4 (0.0052)

Our Metadynamics simulations of aR13aG, bG14aG and bG16aG disaccharides in vacuum predict flexible glycosidic bonds with extensive regions below 4kcal/mol<sup>2</sup> on the resulting PMF surfaces (Figure 15 a, d, e and f, i, j). The plots of unbiased, aqueous phase simulation data presented in Figures 16 and 17 a, d and e, however, indicate that rotation about these bonds is restricted in the central RUs of both the O25a' and O25b' trimers. This restriction is likely due to steric interactions between adjacent monosaccharides.<sup>6,75</sup> The C1, C3, C4 and C6 atoms of  $\alpha$ -D-glucopyranose are all involved in glycosidic bonds in both trimers, the surrounding space is largely occupied and movement of the adjoining monosaccharides is constrained. This explanation is supported by noting that rotation around the aR13aG and bG16aG bonds at the non-reducing ends of the oligosaccharides is not constrained in this way (as the adjacent bG14aG bond is absent) - refer to Figure 18.



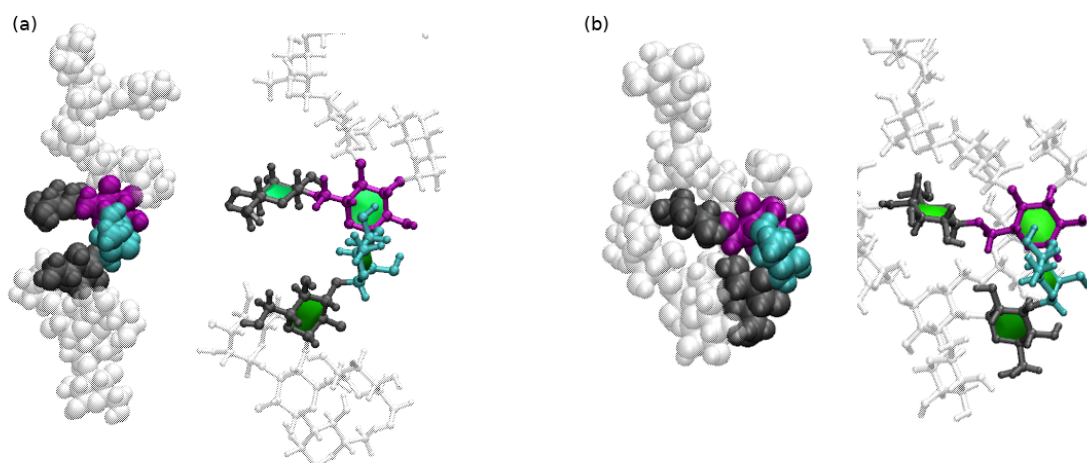
**Figure 18:** The regions around the central  $\alpha$ -D-glucopyranose monosaccharide of the O25a' and O25b' trimers are largely occupied. Representative conformers extracted from our solution phase CHARMM36 trajectories of the (a) O25a' trimer and (b) the O25b' trimer are shown. The central RUs and adjacent, non-reducing  $\beta$ -D-glucopyranose monosaccharides are highlighted and expanded. Possible non-conventional hydrogen bonds stabilizing the aR13bG linkage orientations are shown as black dotted lines. ( $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc in grey,  $\alpha$ -L-Rha in cyan, and  $\alpha$ -L-Fuc in blue)

A stabilizing, non-conventional hydrogen bond where an aliphatic hydrogen atom acts as a donor has been identified between H-C (5) of L-fucopyranose and O (5) of D-galactopyranose in the sialyl Lewis-X pentasaccharide.<sup>76</sup> NMR evidence for this bond in aqueous solution has been reported, as has evidence for hydrogen bonds between the H-C (5) L-fucopyranose atom and the bridging O (4) D-galactopyranose atom of the pentasaccharide.<sup>77</sup> Interestingly, simulations of the Lewis pentasaccharide using GLYCAM06h found that it retained a rigid conformation even when partial charges on aliphatic hydrogens were set as 0 (i.e. without the stabilizing impact of the hydrogen bond).<sup>77</sup> Both our CHARMM36 and GLYCAM06 calculations indicate that the aR13aG bonds of all of the central O25 RUs are constrained around similar dihedral values (Tables 2 and 4). The preferred 3D arrangements of the  $\alpha$ -L-rhamnopyranose and nearby  $\beta$ -D-glucopyranose units (Figure 18) resemble those reported for L-fucopyranose and D-galactopyranose in the sialyl Lewis-X pentasaccharide.<sup>76</sup> The distance and angles between the O25 H-C (5) atom of  $\alpha$ -L-rhamnopyranose and the O (5) and O (4) atoms of  $\beta$ -D-glucopyranose for representative conformers are in reasonable agreement with the range reported (2.6 Å and 134 ° ; 3.4 Å and 160 °) for non-conventional hydrogen bonding as well.<sup>77</sup> Finally, the CHARMM36 plot for the aLRha13aGlc bond in the O25b' trimer (Figure 17 a) shows a small population of conformers at  $\phi \sim -40^\circ$ . The a-

L-rhamnopyranose H-C (5) atom is close (around 2.4 Å away) from the O (5) ring oxygen atom of  $\beta$ -D-glucopyranose in selected conformers with these dihedral angles. This population may result from attractive electrostatic interactions between these atoms in CHARMM36 (where aliphatic hydrogens carry a small positive point charge) that are absent in GLYCAM06.

A cursory VMD H-bonds analysis shows a 25% occupation for the H-C (5) O (5) bond at a cutoff of 3.5 Å & 30 ° for O25b. We do not include an NMR investigation here, but, based on the preceding discussion, it is feasible that non-conventional hydrogen bonds stabilize the orientation of the O25  $\alpha$ -L-rhamnopyranose side chain. Further investigation is required to confirm this.

There are several key differences between the CHARMM36 and GLYCAM06 dihedral angle data for the O25b' RU: GLYCAM06 predicts that the aG13aR bond (Figure 17 c) favors conformers with more negative dihedral angles than CHARMM36; GLYCAM06 predicts a population of conformers in a high energy region around ( $\varphi = -59.1^\circ$ ,  $\Psi = -36.0^\circ$ ) for the aR13bG bond (Figure 17 d); finally, CHARMM36 and GLYCAM06 favor different orientations of the bG16aG linkage (Figure 15 e). These features are due to the collapse of the O25b' trimer into a compressed, metastable state during the GLYCAM06 simulation that does not occur during the CHARMM36 simulation. Refer to Figure 19.



**Figure 19:** Representative conformers extracted from our (a) CHARMM36 and (b) GLYCAM06 O25b' trimer simulation trajectories. Note that the central  $\beta$ -D-glucopyranose unit in (b) is surrounded by other monosaccharides in the collapsed GLYCAM06 conformation. The  $\beta$ -D-glucopyranose unit is not constrained in this way in the CHARMM36 conformer presented in (a). ( $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc in grey,  $\alpha$ -L-Rha in cyan).

In our GLYCAM06 simulation, the dihedral values of the central aG13aR and aR13bG glycosidic linkages of the O25b' trimer are modified as the reducing end of the molecule folds back toward the center. The effect is more pronounced for the aR13bG linkage than it is for aG13aR as it is closer to the tight bend illustrated in Figure 19 b. Indeed, the attractive interactions that appear to be driving the folding of O25b' in GLYCAM06 are enough to move the orientation of the aR13bG bond into a region with an expected vacuum energy of about 7 kcal/mol (with some conformers reaching 9kcal/mol). Our analogous CHARMM36 simulation does not exhibit this feature for the aR13bG linkage and indicates that an overwhelming majority of conformers have dihedral values with energies below 4 kcal/mol for this bond – Figure 17 c.

Our CHARMM36 data favors an orientation of ( $\Psi = -176.5^\circ$ ,  $\omega = -67.5^\circ$ ) for the bG16aG side chain linkage of the O25b' trimer whereas our GLYCAM06 data favors a notably different arrangement around ( $\Psi = 171.5^\circ$ ,  $\omega = 157.5^\circ$ ). Again, review of the low energy regions of the CHARMM36 and GLYCAM06 PMF surfaces (Figure 15 e) for this bond does not provide a clear explanation for this large difference and intramolecular interactions appear to be important. The conformers illustrated in Figure 19 provide a rationale: The bG16aG bond is held in position by steric interactions with its many surrounding monosaccharides in the collapsed conformation predicted by GLYCAM06 (Figure 19 b). These constraints are not present in our CHARMM36 simulation (which favors more extended conformers for the O25b' trimer) and so the bG16aG bond adopts an orientation close to that predicted by CHARMM36 for the analogous bond in O25a'. We have not investigated hydrogen bonding involving the  $\beta$ -D-glucopyranose side chain in the collapsed conformation predicted by GLYCAM06 but there are certainly candidate arrangements available. Note that non-conventional hydrogen bonding is not the driver of the collapse of the O25b' trimer in our GLYCAM06 simulation as aliphatic hydrocarbons do not carry a partial charge in this force field.<sup>77</sup> Other types of inter- and intramolecular interactions are involved.

Despite some differences, the preferred orientations of the glycosidic bonds of the central O25a' RUs predicted by our CHARMM36 and GLYCAM06 simulations are in reasonable agreement. The CHARMM36 data indicates some curved conformers that are absent in GLYCAM06. These conformers have central aF13bG bonds with ( $\phi \sim 35^\circ$ ,  $\Psi \sim 169.5^\circ$ ) - refer to Figure 4 c. The reason for the presence of these arrangements in CHARMM36 and not in GLYCAM06 is not obvious but the deviation may account for the small difference in the preferred end-to-end distance for the

O25a' trimer predicted by the two force fields. Subtle differences in how intermolecular interactions with the solvent are handled by CHARMM36 and GLYCAM06 may be a factor here – electrostatic interactions are handled differently by these two force fields.<sup>22</sup>

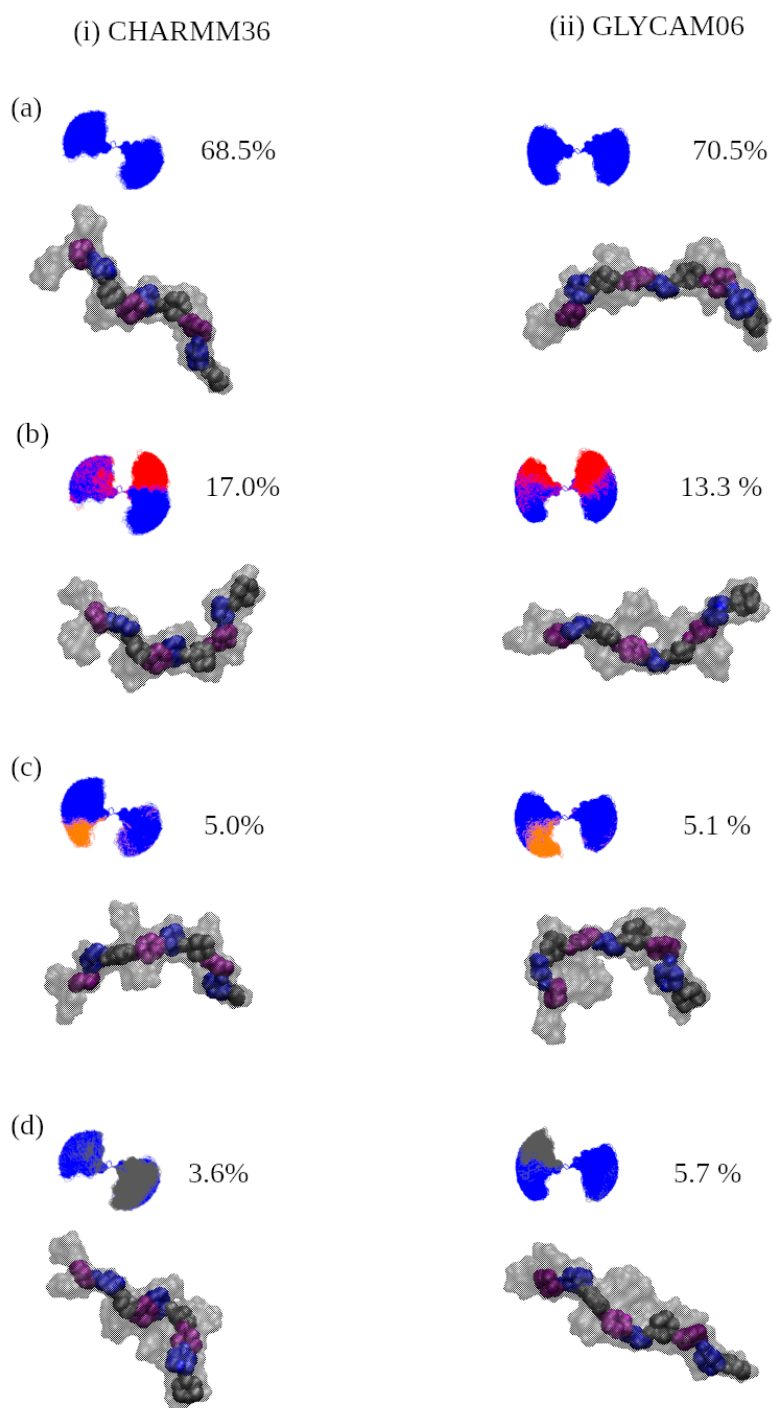
### 7.1.3 RMSD cluster analysis of O25a' and O25b' trimers

Root Mean Square Deviation (RMSD) analysis provides a measure of the overall variation of a population of conformers from one another and is a useful metric for grouping related arrangements into clusters. RMSD clustering of 3 RU oligosaccharide extensions of O25a' and O25b' is presented here as an extension of the discussion covered in Sections 7.1.1 and 7.1.2 and to provide an additional view of the conformations predicted for the trimers and their relative weightings.

Each oligosaccharide extensions is aligned to the central monosaccharide of their central RUs (fucose in O25a' and rhamnose in O25b'). These units were selected as they are conveniently located at the center of each trimer and we have established that their glycosidic bonds are important contributing factors to the overall conformations of the molecules. We have disregarded the first 100 ns of simulation time for equilibration and focus our analysis on the last 100 ns sampling conformers every 2.5 ps.

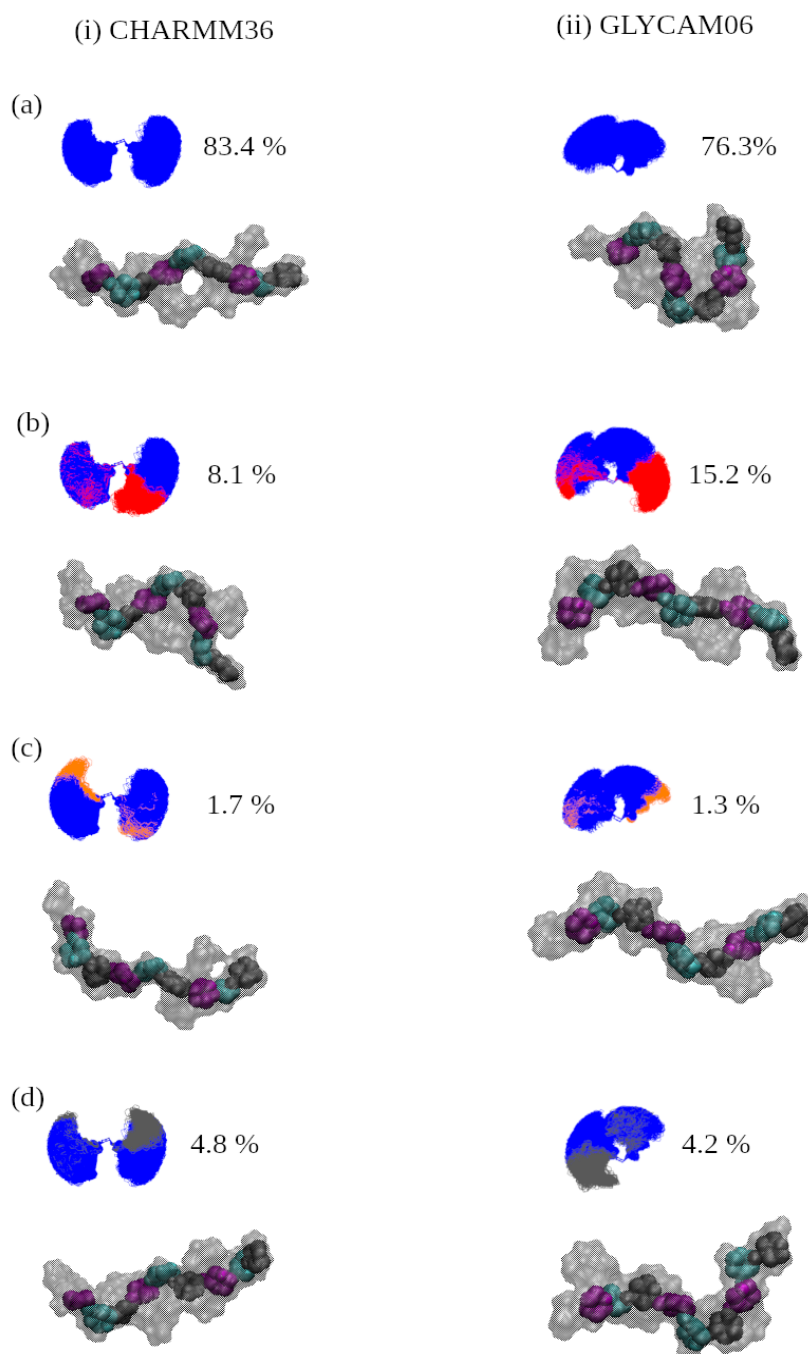
RMSD clusters, weightings and representative conformers of the O25a trimer from our CHARMM36 and GLYCAM06 simulation data are presented in Figure 20. The overwhelming majority of conformers in both simulations exhibit curved, linear arrangements that are largely consistent with the level of extension data presented in Section 7.1.1. The reducing and non-reducing end  $\alpha$ -D-glucose monosaccharides remain separated in these conformations with their distance of separation,  $r$ , remaining between 25 and 30 Å. GLYCAM06 appears to favor curved O25a' conformers somewhat more than CHARMM36 which accounts for the slightly lower GLYCAM06  $r$  value predicted for this molecule.

RMSD clusters for the O25b' trimer are presented in Figure 21. The compact, metastable conformation in the GLYCAM06 simulation is clear and includes the overwhelming majority of conformers (with a weighting of 76.3 %). As expected from the discussion presented in Sections 7.1.1 and 7.1.2, CHARMM36 does not exhibit this metastable conformation. Most CHARMM36 conformers are curved, linear or loose right handed helical arrangements. CHARMM36 predicts that O- and N-acetyl groups on backbone O25b units (which are absent in O25b' here) have a major stabilising effect on the helical conformation. This is discussed further in Section 7.2.



**Figure 20:** RMSD clustering of a 3 RU O25a' oligosaccharide extensions aligned to the central  $\alpha$ LFuc monosaccharide for (i) CHARMM36 and (ii) GLYCAM06. 4 clusters with a cutoff value of 5 were selected grouping 94.1 % and 94.6 % of conformers respectively. All views are orientated with the non-reducing end on the left. Clustering is based on the arrangement of the highlighted backbone monosaccharides. (a) Shows the arrangement, weighting and a representative conformer for the largest cluster. (b), (c), (d) Show the arrangements, weightings and conformers of the other 3 clusters in relation to (a).  $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc in grey,  $\alpha$ -L-Fuc in blue.





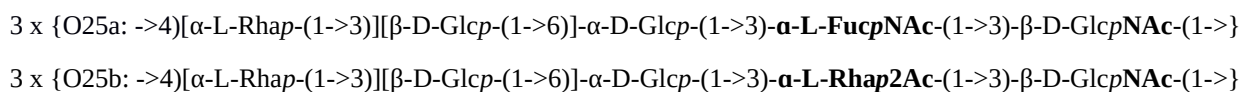
**Figure 21:** RMSD clustering of a 3 RU O25b' oligosaccharide extensions aligned to the central aLRha monosaccharide for (i) CHARMM36 and (ii) GLYCAM06. 4 clusters with a cutoff value of 5 were selected grouping 98.0 % and 97.0 % of conformers respectively. All views are orientated with the non-reducing end on the left. Clustering is based on the arrangement of the highlighted backbone monosaccharides. (a) Shows the arrangement, weighting and a representative conformer for the largest cluster. (b), (c), (d) Show the arrangements, weightings and conformers of the other 3 clusters in relation to (a). a-D-Glc in purple, β-D-Glc in grey, α-L-Rha in cyan.

#### 7.1.4 Concluding remarks

Our comparative investigation of the constituent glycosidic bonds of the central RUs of the O25a' and O25b' trimers provides several insights: First, despite some differences for the aG13aF and aF13bG linkages, the dihedral angles of the constituent glycosidic bonds of the central O25a' RU predicted by our CHARMM36 and GLYCAM06 simulations are in agreement. Second, our CHARMM36 and GLYCAM06 simulations predict markedly different dihedral values for the main-chain aR13bG bond and side-chain bG16aG bond of the central RU of the O25b' trimer. A rationale for these differences is the collapse of the molecule into a compressed, metastable state during the GLYCAM06 simulation. The collapsed antigen conformation forces a bend in the main oligosaccharide chain and stabilizes an anti-anti orientation of the  $\beta$ -D-glucopyranose side chain in the GLYCAM06 simulation. Finally, there is a high degree of conformance in the dihedral values predicted for the aR13aG side chains of all of the O25 RUs in both our CHARMM36 and GLYCAM06 simulations. The predicted arrangement of these groups is stabilized by a hydrogen bond between the side-chain  $\alpha$ -L-rhamnopyranose C4 hydroxyl group and the main-chain  $\beta$ -D-glucopyranose C6 hydroxyl group. The orientation of these units is similar to a fragment of the sialyl Lewis-X pentasaccharide and may involve a non-conventional hydrogen bond.<sup>76</sup>

## 7.2 Comparison of O25a and O25b conformations

Having assessed the conformations of O25a' and O25b' oligosaccharide extensions using both CHARMM36 and GLYCAM06, we now look to enhance our understanding of the O25 antigens by considering the effects of their O- and N-acetyl groups on their conformations:

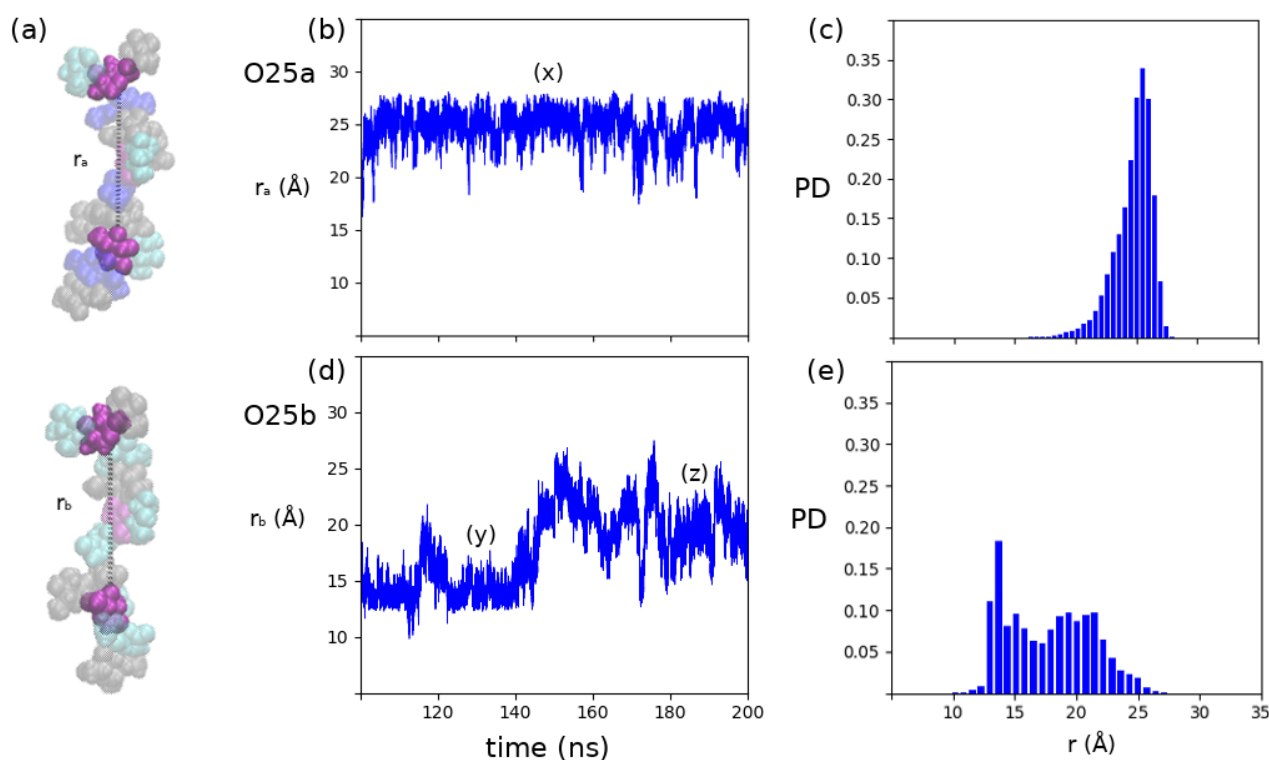


The same approach that was applied to our analysis of O25a' and O25b' is applied here. First, we review the level of extension of 3 RU oligosaccharide extensions of O25a and O25b in Section 7.2.1. Thereafter, we proceed to an analysis of the dihedral angles of the central RUs of each trimer in Section 7.2.2. Only CHARMM36 simulation data is available for the acetylated oligosaccharides. Initial dihedral angles for the starting structures of the O25a and O25b oligosaccharide extensions were informed by the disaccharide PMF surfaces presented in Figure 22.



### 7.2.1 Preferred level of extension of O25a and O25b trimers

Once again, the distance between the ring oxygen atoms of the  $\alpha$ -D-glucopyranose monosaccharides nearest to the reducing and non-reducing ends of each oligosaccharide extension are used as a measure of their preferred levels of extension. The values of  $r_a$  and  $r_b$  represent this distance for the O25a and O25b trimers respectively. We have disregarded the first 100 ns of simulation time for equilibration leaving 100 ns for analysis. Plots of  $r_a$  and  $r_b$  are presented in Figure 22.



**Figure 22:** (a) Models of the O25a and O25b trimers showing our selected measure of extension,  $r$ .  $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc and  $\beta$ -D-GlcNAc in grey,  $\alpha$ -L-Rha2Ac in cyan,  $\alpha$ -L-FucNAc in blue. (left) Time series end-to-end distance plots of our (b) O25a and (d) O25b trimers from CHARMM36 simulation data. (x), (y) and (z) represent the positions of the conformers illustrated in Figure 19. (right) Probability density plots of the end-to-end distance of the (c) O25a and (e) O25b trimers using CHARMM36 simulation data.

The time series end-to-end distance plot of  $r_a$  (Figure 22 b) shows rapid transitions between 16.1 Å and 28.2 Å with an average of 24.7 Å (standard deviation of 1.5 Å). These values are comparable to those calculated from CHARMM36 simulation data for the unacetylated O25a' oligosaccharide extension (15.1 Å to 28.1 Å with an average of 24.3 Å).

The probability density plot of  $r_a$  (Figure 22 c) exhibits a distinct peak at about 25.5 Å with a height of 0.34. The maximum of this curve is more prominent than that presented for O25a' (Figure 9 c), the gradient of the leading end is steeper and the curve is shifted toward larger values of  $r$ .

Overall, O- and N-acetylation appears to have a small extending effect on the O25a trimer compared to our O25a' analogue that lacks these groups. Simulations of longer O25a polysaccharides comprised of 6 or 12 RUs may be of interest to confirm the impact of O- and N-acetylation on O-antigen conformation *in vivo*.

The time series end-to-end distance plot of  $r_b$  (Figure 22 d) appears different to that presented for our O25b' trimer (Figure 9 b). The  $r_b$  plot indicates transitions between 9.8 Å and 27.5 Å with an average of 17.7 Å (standard deviation of 3.4 Å). These values indicate that the acetylated O25b trimer is more compressed than its un-acetylated O25b' analogue (with transitions between 13.2 Å and 28.1 Å, average 22.6 Å, standard deviation 2.0 Å).

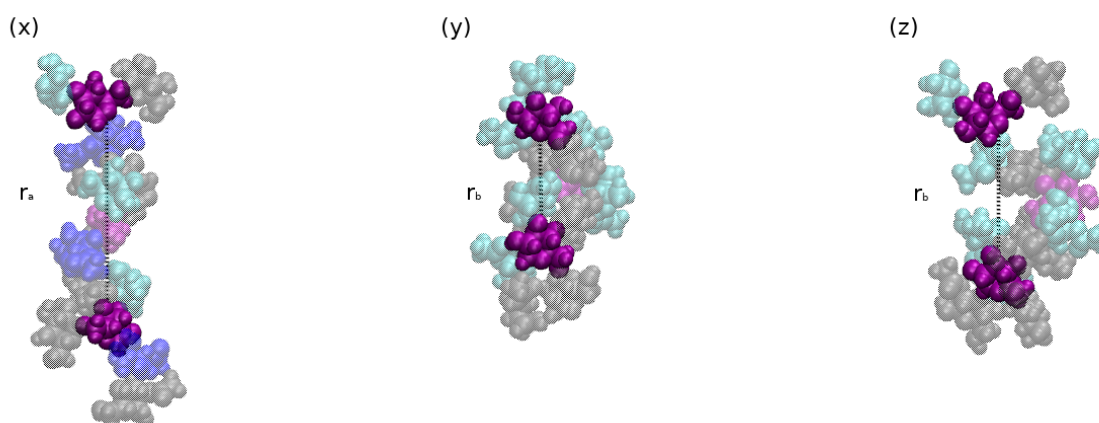
The probability density plot of  $r_b$  (Figure 22 e) presents a global maximum of 0.18 at 13.7 Å which indicates a much shorter oligosaccharide than that calculated for the unacetylated O25b' trimer from the CHARMM36 simulation data (0.22 at 22.6 Å - Figure 9 c). Interestingly, the most probable  $r_b$  value is similar to the most probable  $r$  value for O25b' (15.5 Å) calculated from GLYCAM06 data in Section 7.1.1. This suggests some form of compressed conformation of the O25b trimer. A secondary, broad peak is visible at around 21 Å in Figure 22 e indicating that the molecule is not locked into a collapsed conformation in the CHARMM36 simulation. Finally, the O25b probability density curve (Figure 22 e) is shifted towards lower  $r_b$  values than that for O25b' (Figure 9 c).

Our CHARMM36 data indicates that the presence of O- and N-acetyl groups has a significant compressive effect on the O25b trimer compared to our O25b' analogue that lacks these groups. Based on this difference, O- and N-acetylation is likely to have an important effect on the conformations of longer O25b polysaccharides and studies of 6 and 12 RU oligosaccharide extensions are of interest.

**Table 3:** The preferred level of extension of the O25a', O25b', O25a and O25b trimers presented as the end-to-end distance between the ring oxygen atoms of their reducing and non-reducing  $\alpha$ -D-glucose units ( $r$ ). Minimum, maximum, average and most probable values calculated from CHARMM36 and GLYCAM06 simulation data are indicated. Standard deviations (for averages) and probability density peak heights (for most probable values) are bracketed. Values for the acetylated O25a and O25b trimers are highlighted in bold.

	CHARMM36				GLYCAM06				Delta
	Min. $r$ (Å)	Max. $r$ (Å)	Avg. $r$ (Å)	Prob. $r$ (Å)	Min. $r$ (Å)	Max. $r$ (Å)	Avg. $r$ (Å)	Prob. $r$ (Å)	
O25a'	15.1	28.1	24.3 (1.6)	25.2 (0.27)	15.3	29.2	25.1 (1.8)	26.1 (0.26)	0.9
<b>O25a</b>	<b>16.1</b>	<b>28.2</b>	<b>24.7 (1.5)</b>	<b>25.5 (0.34)</b>	-	-	-	-	-
O25b'	13.2	28.1	22.6 (2.0)	23.5 (0.22)	9.97	28.4	18.0 (3.8)	15.5 (0.29)	8
<b>O25b</b>	<b>9.8</b>	<b>27.5</b>	<b>17.7 (3.4)</b>	<b>13.7 (0.18)</b>	-	-	-	-	-

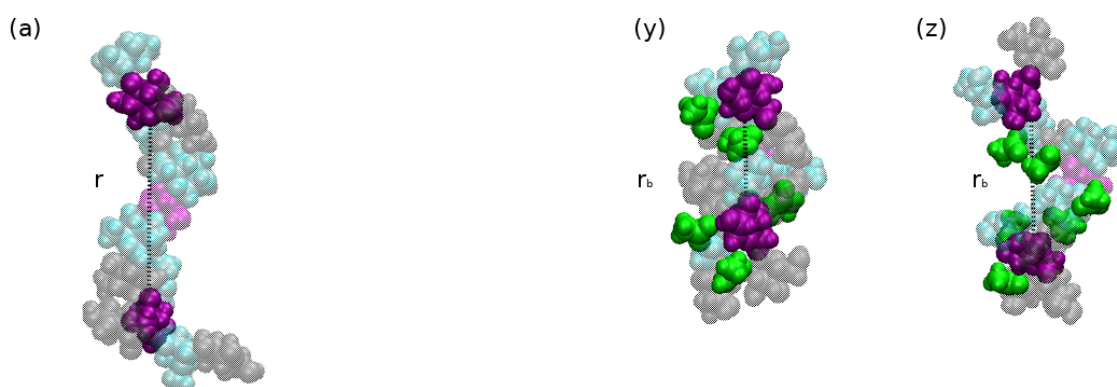
Representative conformers that correspond to the points labeled (x), (y) and (z) marked on the time series plots in Figure 22 b and d are presented in Figure 23 x y and z respectively. Note that our CHARMM36 simulation data indicates a fairly high level of flexibility of the O25b trimer, with conformers ranging from collapsed (Figure 23 y) to somewhat more extended arrangements (Figure 23 z). Overall, however, the O25b trimer is predicted to have a much more compressed conformation than its O25a analogue which appears to favor a more rigid, extended, linear conformation (Figure 23 x).



**Figure 23:** Representative conformers, showing our selected measure of extension, extracted from the trajectories of CHARMM36 simulations for: (x) O25a where  $r_a = 22.4$  Å at 145.997 ns, (y)

O25b where  $r_b = 14.3 \text{ \AA}$  at 130.561 ns and (z) O25b where  $r_b = 19.8 \text{ \AA}$  at 187.215 ns. ( $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc and  $\beta$ -D-GlcNAc in grey,  $\alpha$ -L-Rha and  $\alpha$ -L-Rha2Ac in cyan,  $\alpha$ -L-FucNAc in blue)

The large difference in the predicted levels of extension of the O25b' and O25b trimers is attributed to the presence of O- and N-acetyl groups on the O25b oligosaccharide model that were omitted from the O25b' model. The arrangement, inter- and intermolecular interactions of these groups are thus likely to be important contributors to the conformation of the O25b antigen. Representative O25b' and O25b conformers with their O- and N-acetyl groups highlighted are illustrated in Figure 20. Note the substantial difference in preferred  $r$  values predicted for O25b' and O25b (Table 3).



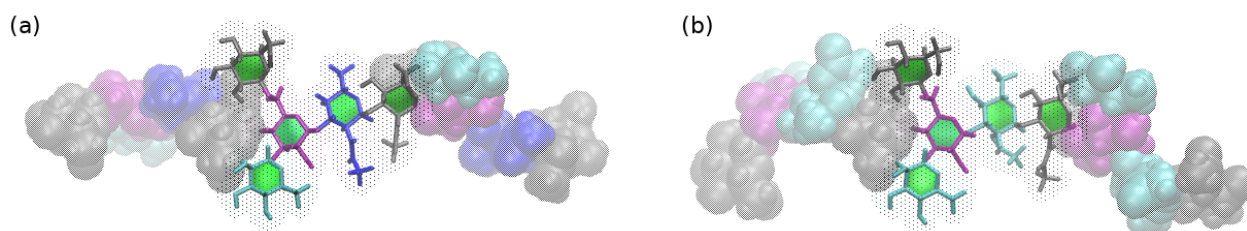
**Figure 24:** Representative conformers extracted from the trajectories of CHARMM36 simulations for: (a) O25b' where  $r = 23.6 \text{ \AA}$  at 145.058 ns, (y) O25b where  $r_b = 14.3 \text{ \AA}$  at 130.561 ns and (z) O25b where  $r_b = 19.8 \text{ \AA}$  at 187.215 ns. O- and N-acetyl groups are highlighted in green. ( $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc and  $\beta$ -D-GlcNAc in grey,  $\alpha$ -L-Rha and  $\alpha$ -L-Rha2Ac in cyan,  $\alpha$ -L-FucNAc in blue)

In summary, our comparative analysis of aqueous phase, CHARMM36 simulation data of the level of extension of O25a and O25b trimers provides several insights: First, O- and N-acetylation of the O25a trimer is predicted to have only a small effect on the level of extension of the oligosaccharide. O- and N-acetylation does appear to make O25a somewhat more rigid than its unacetylated O25a' analogue however. Second, the O25b trimer strongly favors compressed conformations (with smaller  $r$  values) over more the extended conformations favored by the O25b' trimer. This preference is due to the presence O- and N-acetyl groups in our O25b model that were omitted from the O25b' model. The positions and interactions of these groups are thus important determinants of the conformation of the O25b antigen. Finally, the O25b trimer favors more much compressed conformations than the O25a trimer which orientates into a fairly rigid, extended arrangement. The

predicted difference between O25a and O25b is large and significant conformational differences between these antigens are considered to be likely *in vivo*.

### 7.2.2 Preferred glycosidic bond dihedrals for O25a and O25b

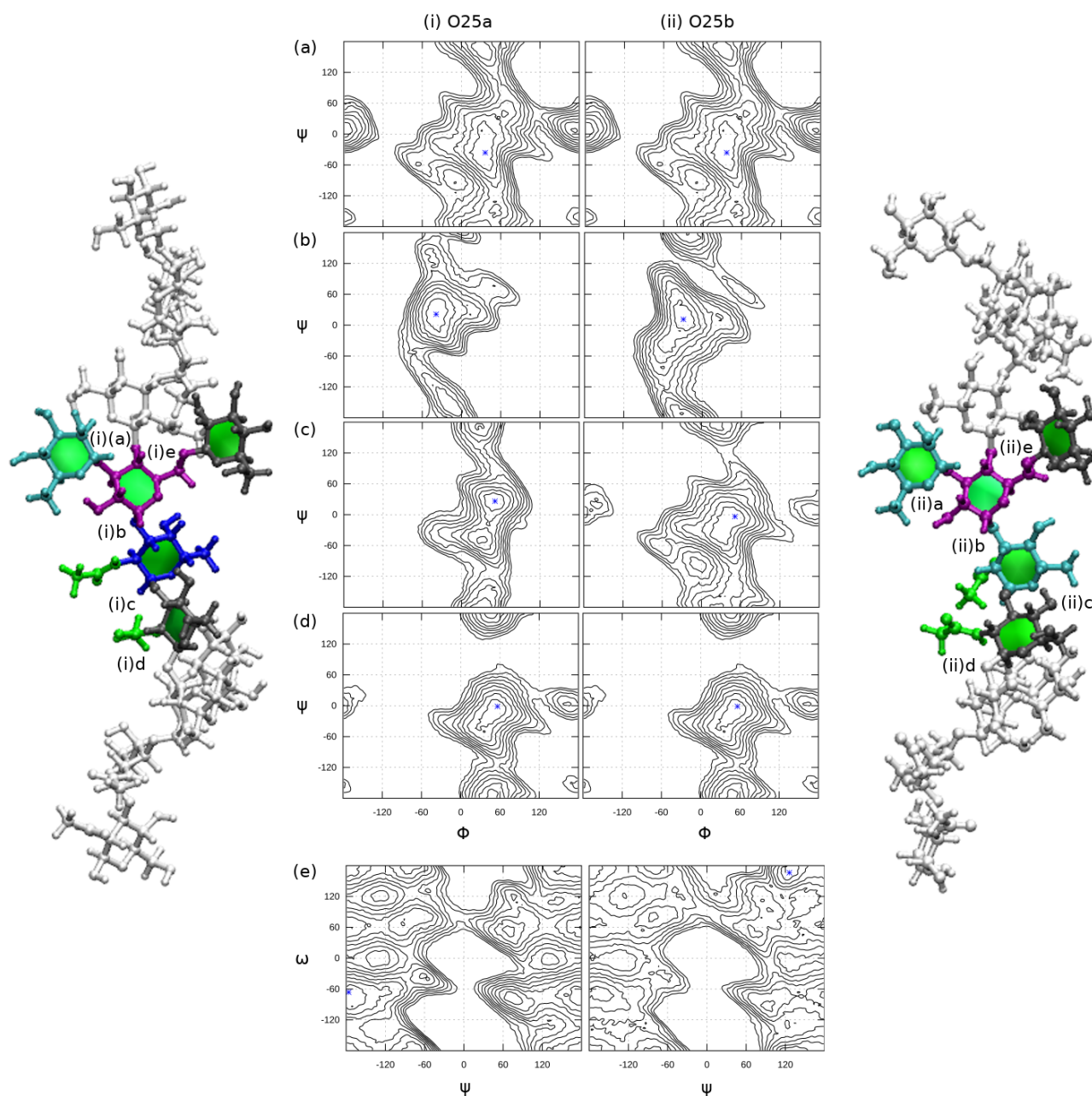
Having reviewed the effects of O- and N-acetylation on the preferred level of extension of O25a and O25b trimers in Section 7.2.1, we now turn our attention to the dihedral angles of the constituent glycosidic bonds of these two molecules. The objectives here are to further rationalise the large difference in the levels of extension predicted for the O25a and O25b trimers and to explore the inter- and intramolecular interactions of their respective O- and N-acetyl groups. These interactions may provide a rationale for the important effect that O- and N-acetylation appears to have on the predicted conformation of the O25b polysaccharide. The same approach that was applied to our analysis of the dihedral values of the glycosidic bonds of O25a' and O25b' is applied here. Once again, future references to the O25a and O25b RUs refer to the central pentasaccharide illustrated in Figure 25.



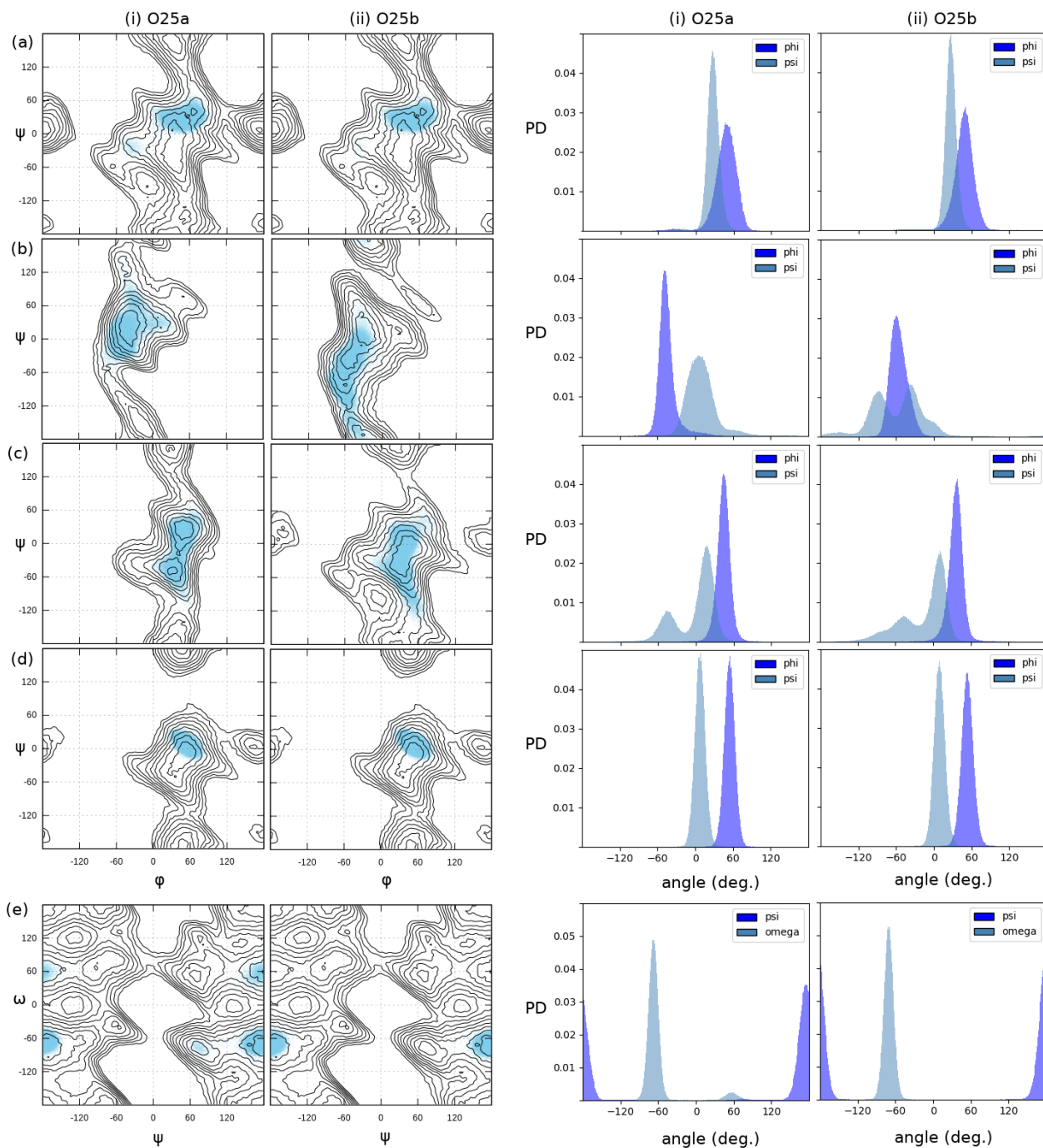
**Figure 25:** Minimized oligosaccharide models highlighting the glycosidic bonds of the central RUs of (a) O25a and (b) O25b. Note the presence of acetyl and N-acetyl groups. ( $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc and  $\beta$ -D-GlcNAc in grey,  $\alpha$ -L-Rha and  $\alpha$ -L-Rha2Ac in cyan, and  $\alpha$ -L-FucNAc in blue)

PMF energy surfaces of vacuum, Metadynamics simulation data for disaccharides that represent the constituent glycosidic bonds of the O25a and O25b RUs are presented in Figure 26. The low energy regions of these plots were used to inform our selection of glycosidic bond dihedral values for the starting structures of our O25a and O25b oligosaccharide extensions. A more extensive analysis of these PMF surfaces is presented in Appendix A. Only CHARMM36 data is presented here. The dihedral values of the glycosidic bonds of the starting structures were set in close agreement with the respective PMF energy minima. Data from solution simulations of the O25a and O25b trimers is presented in Figure 27 and summarized in Table 4. Once again, we have disregarded the first 100 ns of simulation time for equilibration leaving 100 ns for analysis.





**Figure 26:** PMF surfaces plotted from CHARMM36 vacuum simulation data of representative disaccharides for the glycosidic bonds of (i) O25a and (ii) O25b. Plots are presented for: O25a (i) (a) aR13aG (i)(b) aG13aFNAc (i)(c) aFNAc13bGNac (i)(d) bGNac14aG (i)(e) bG16aG, O25b' (ii) (a) aR13aG (ii)(b) aG13aRAc (ii)(c) aRAc13bGNac (ii)(d) bGNac14aG (ii)(e) bG16aG. Contour lines are plotted in increments of 1kcal/mol up to a cutoff of 9kcal/mol. Global minimum energy values are presented as blue stars. Static models of our starting structures for O25a (left) and O25b (right) trimers are presented with O- and N-acetyl groups highlighted in green. The central RU of each trimer is highlighted. ( $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc and  $\beta$ -D-GlcNAc in grey,  $\alpha$ -L-Rha and  $\alpha$ -L-Rha2Ac in cyan,  $\alpha$ -L-FucNAc in blue)



**Figure 27:** (left) Scatter plots and (right) probability density plots for the dihedral angle populations of the central RUs of the (i) O25a and (ii) O25b trimers. O25a data for the (i)(a) aR13aG (i)(b) aG13aFNAc (i)(c) aFNAc13bGNac (i)(d) bGNac14aG and (i)(e) bG16aG bonds is presented. O25b data for the (ii)(a) aR13aG (ii)(b) aG13aRAc (ii)(c) aRAc13bGNac (ii)(d) bGNac14aG and (ii)(e) bG16aG bonds is presented. Contour lines are in increments of 1kcal/mol up to a cutoff of 9 kcal/mol.

**Table 4:** Preferred dihedral angle values calculated from probability density plots of the constituent glycosidic bonds of the O25a and O25b RUs (Figures 27). Probability density peak heights are indicated in brackets next to their respective dihedral angle values.

RU	Glycosidic Bond	O25a dihedral angles		
		$\varphi / ^\circ$	$\Psi / ^\circ$	$\omega / ^\circ$
O25a	aR13aG	46.9 (0.027)	25.9 (0.046)	-
<b>O25a</b>	<b>aG13aFNAc</b>	<b>-48.6 (0.042)</b>	<b>5.9 (0.021)</b>	-
<b>O25a</b>	<b>aFNAc13bGNAc</b>	<b>45.7 (0.043)</b>	<b>17.1 (0.024)</b> <b>-47.9 (0.0079)</b>	-
O25a	bGNAc14aG	54.5 (0.049)	6.5 (0.049)	-
O25a	bG16aG	-	176.5 (0.036)	-68.8 (0.049) 56.7 (0.0025)
O25b	aR13aG	50.5 (0.031)	27.1 (0.050)	-
<b>O25b</b>	<b>aG13aRAc</b>	<b>-60.4 (0.032)</b>	<b>-36.5 (0.013)</b> <b>-87.5 (0.011)</b>	-
<b>O25b</b>	<b>aRAc13bGNAc</b>	<b>37.2 (0.042)</b>	<b>11.0 (0.023)</b> <b>-44.22 (0.0066)</b>	-
O25b	bGNAc14aG	53.0 (0.044)	8.4 (0.048)	-
O25b	bG16aG	-	178.5 (0.044)	-69.8 (0.055)

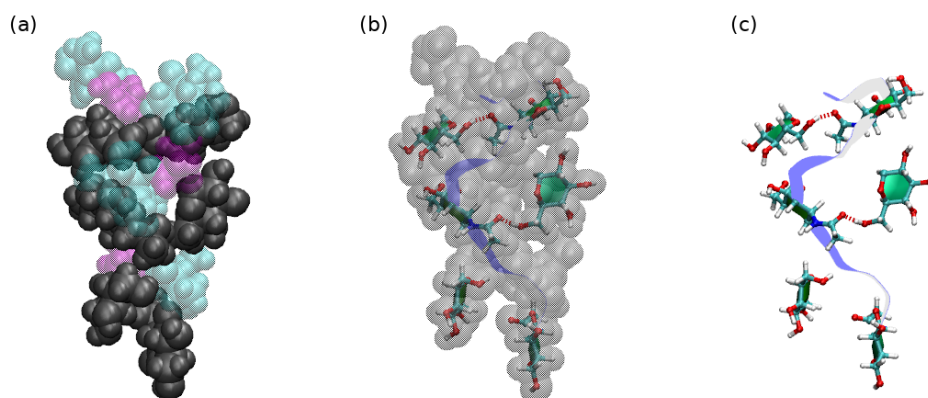
The PMF surfaces presented in Figure 26 provide some insight into the reasons for the differences in the preferred levels of extension of the O25a and O25b trimers noted in Section 7.2.1. The plot for the aG13aRAc bond of the O25b RU (Figure 26 b ii) has an extended, low energy valley connecting the region around its global energy minimum to a local syn-anti energy minimum with  $\Psi \sim -180^\circ$ . This bond is, therefore, free to rotate about a wide range of negative  $\Psi$  values. The PMF plot for the corresponding aG13aFNAc bond in the O25a RU (Figure 26 b i) exhibits a higher energy (about 8 kcal/mol) path for transitions to regions with  $\Psi < -45^\circ$ . Rotation of the aG13aFNAc bond into these orientations is, therefore, constrained by the anomeric effect and steric effects between the bridged monosaccharides.<sup>78</sup> This is borne out in the unbiased simulation data for the O25a and O25b trimers presented in Figure 25 b i and ii: O25a conformers that have aG13aFNAc linkages with  $\Psi < -35^\circ$  are unlikely. This constraint is not present for the aG13aRAc



bond of the O25b RU where large populations of conformers with  $\Psi = -36.5^\circ$  and  $\Psi = -87.5^\circ$  are predicted.

Review of the PMF plot for the unacetylated aG13aR linkage (Figure 15 g i) reveals that 2 O-acetylation of the  $\alpha$ -L-rhamnopyranose unit is responsible for the elongated low energy region noted in Figure 27 b ii. We postulate that 2 O-acetylation of main-chain  $\alpha$ -L-rhamnopyranose units in O25b facilitates rotation about its aG13aRAc bonds into orientations that are constrained in the corresponding aG13aR and aG13aFNAc bonds of O25b' and O25a. This accounts, in part, for the comparably flexible nature of the O25b trimer and its tendency to adopt compressed conformations relative to O25a in solution. Intramolecular interactions also play a key role in the preferred conformation of the O25b trimer, as follows.

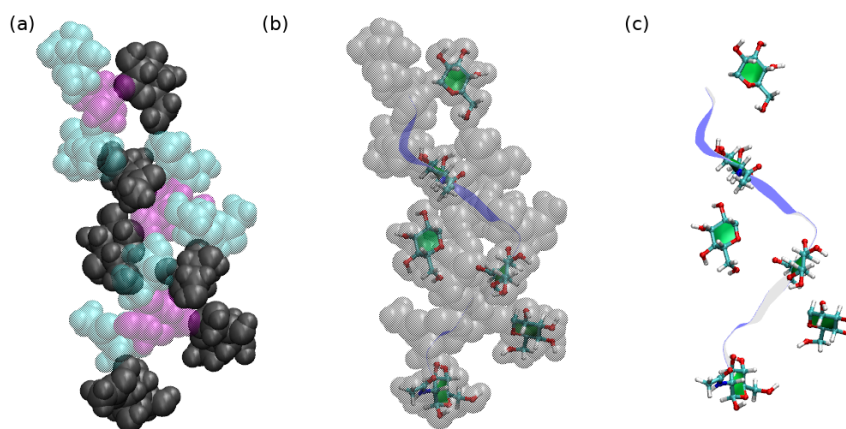
Our end-to-end distance analysis of O25b (Section 7.2.1) suggests the presence of two broad populations of conformers: one compressed group with an end-to-end distance ( $r_b$ ) of around 13.7 Å and a second, more extended group with  $r_b$  values ranging around 21 Å. The prominent  $r_b$  probability density peak at 13.7 Å in Figure 22 e indicates a strongly favored O25b conformation with this property. The glycosidic bonds of the O25b RU should be ordered beyond the constraints of the low energy regions of their PMF surfaces in this conformation. A compressed O25b conformer is illustrated in Figure 28:



**Figure 28:** (a, b and c) Representations of a compressed, 3 RU O25b conformer extracted from the simulation trajectory at 124.020 ns,  $r_b = 13.46$  Å, aG13aRAc dihedral values of ( $\phi = -62.8^\circ$ ;  $\Psi = -88.1^\circ$ ) and aRAc13aGNAc dihedral values of ( $\phi = 36.8^\circ$ ;  $\Psi = 9.7^\circ$ ).  $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc and  $\beta$ -D-GlcNAc in grey,  $\alpha$ -L-Rha and  $\alpha$ -L-RhaAc in cyan. Key hydrogen bonds highlighted in red.

Analysis of the conformer presented in Figure 28 indicates that the N-acetyl-D-glucosamine monosaccharides of the non-reducing and central o25b RUs tend to orientate so that their N-acetyl groups are tilted towards the vertical axis of the molecule when it is in a compressed conformation. This arrangement facilitates hydrogen bonding between the carbonyl oxygen atoms of the main-chain N-acetyl-D-glucosamine monosaccharides and the primary hydroxyl groups of the  $\beta$ -D-glucopyranose side chains. Hydrogen bond analysis (cutoff of 3.5 Å, 30 ° - Appendix H) shows occupations of above 40 % for this interaction in these RUs. Several other hydrogen bonds (with occupations of between 10 % and 19 %) appear to stabilize this compressed, helical conformation as well. aG13aRAc, aRAc13bGNAC, bGNAC14aDGlc and bG16aG dihedral angles in the region of ( $\phi = -62.8^\circ$ ;  $\Psi = -88.1^\circ$ ), ( $\phi = 36.8^\circ$ ;  $\Psi = 9.7^\circ$ ), ( $\phi = 42.3^\circ$ ;  $\Psi = 20.4^\circ$ ) and ( $\Psi = 178.5^\circ$ ;  $\omega = -69.8^\circ$ ) in the central O25b RU are associated with conformers that exhibit this arrangement as per the plots presented in Figure 27 (ii)(b), (ii)(c), (ii)(d) and (ii)(e) respectively.

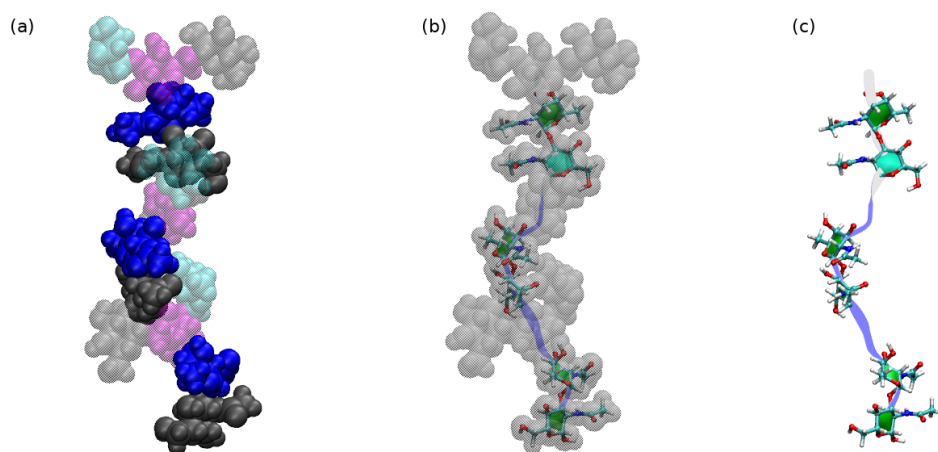
Hydrogen bonds between the N-acetyl-D-glucosamine and  $\beta$ -D-glucopyranose monosaccharides of the terminal, reducing O25b RU do not occur. Instead, the acetylated rhamnose monosaccharide tends to fold towards the non-reducing end of the molecule and its carbonyl group accepts a hydrogen bond from the C2 hydroxyl group of the  $\alpha$ -D-glucopyranose unit of this RU (occupation of 32%). This presents a good departure point for a discussion of the population of more extended O25b conformers. A representative, extended conformer is presented in Figure 29 below:



**Figure 29:** (a, b and c) Representations of an extended conformer extracted from the trajectory of the O25b trimer. Extracted at 195.178 ns,  $r_b = 20.95$  Å, aG13aRAc dihedrals of ( $\phi = -56.42^\circ$ ;  $\Psi = -37.4^\circ$ ) and aRAc13aGNAC dihedral values of ( $\phi = 33.12^\circ$ ;  $\Psi = -43.1^\circ$ ).  $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc and  $\beta$ -D-GlcNAc in grey,  $\alpha$ -L-Rha and  $\alpha$ -L-RhaAc in cyan.

As can be seen in Figure 29, hydrogen bonds between the carbonyl groups of N-acetyl-D-glucosamine monosaccharides and the primary hydroxyl groups of  $\beta$ -D-glucopyranose side chains are absent in all O25b RUs when the trimer is in an extended conformation. Instead, a set of antagonistic hydrogen bonds form between the C2 hydroxyl groups of  $\alpha$ -D-glucopyranose and the carbonyl groups of O-acetylated rhamnose monosaccharides in the central and reducing RUs (occupations of 30.2 % and 15.7 % respectively). These interactions are complemented by a hydrogen bond between the C4 hydroxyl group of the  $\alpha$ -L-rhamnopyranose and  $\beta$ -D-glucopyranose side chains of the central and non-reducing RUs (occupation 10.15 %). Collectively, these interactions draw the  $\beta$ -D-glucopyranose side chains away from the main-chain, N-acetyl-D-glucosamine monosaccharides and result in a more extended conformation. aG13aRAc, aRAc13bGNAc, bGNAc14aDGlc and bG16aG dihedral angles in the region of ( $\varphi = -62.8^\circ$ ;  $\Psi = -37.4^\circ$ ), ( $\varphi = 33.1^\circ$ ;  $\Psi = -43.1^\circ$ ), ( $\varphi = 63.4^\circ$ ;  $\Psi = -0.3^\circ$ ) and ( $\Psi = 178.5^\circ$ ;  $\omega = -69.8^\circ$ ) in the central O25b RU are associated with conformers that exhibit this arrangement – refer to Figure 27 (ii)(b), (ii)(c), (ii)(d) and (ii)(e) and the data presented in Table 4.

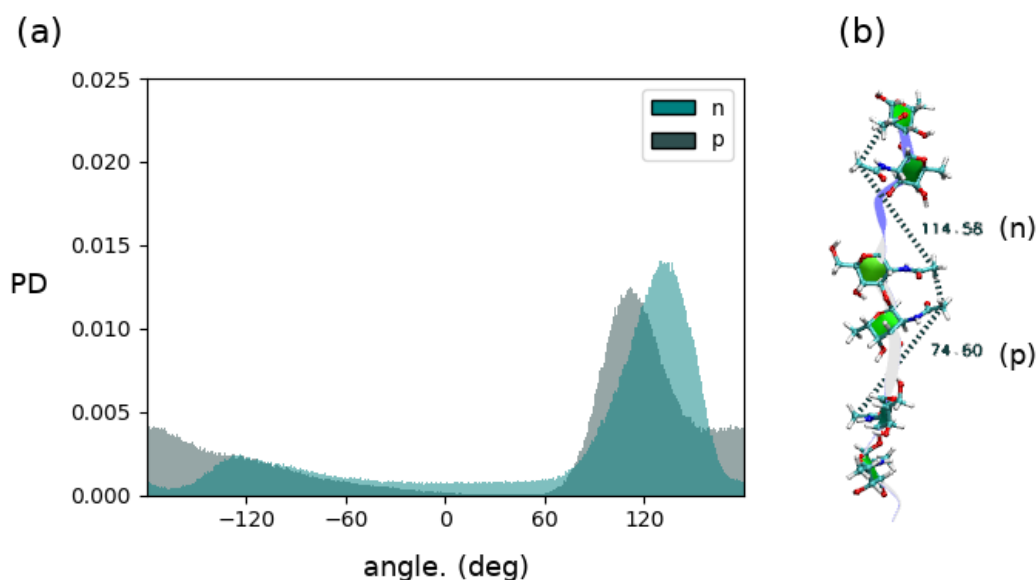
Our end-to-end distance analysis of the O25a trimer indicates that it prefers a more extended, rigid conformation than the O25b trimer. An extended O25a conformer, extracted from the position labeled (x) in Figure 18 (b), is presented in Figure 22.



**Figure 30:** (a, b and c) Representations of a conformer extracted from the trajectory of the O25a trimer. Extracted at 145.997 ns,  $r_a = 22.4 \text{ \AA}$ , aG13aRAc dihedrals of ( $\varphi = -56.42^\circ$ ;  $\Psi = -37.4^\circ$ ) and aRAc13aGNAc dihedral values of ( $\varphi = 33.12^\circ$ ;  $\Psi = -43.1^\circ$ ).  $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc and  $\beta$ -D-GlcNAc in grey,  $\alpha$ -L-Rha and  $\alpha$ -L-RhaAc in cyan,  $\alpha$ -L-FucNAc in blue.

The PMF surfaces of the main-chain, aDG13aFNAc and aFNAc13bGNAc glycosidic linkages of O25a show smaller low energy regions than those for the corresponding O25b bonds – see Figure 27 (i)(b) and (i)(c). Rotation about these bonds is thus more constrained in O25a than it is in O25b. This is in agreement with a more rigid O25a oligosaccharide.

The positions of O- and N-acetyl groups of the main-chain rhamnopyranose and glucopyranose monosaccharides are in a staggered arrangement about the vertical axis of the O25a molecule presented in Figure 30. An analysis of a pair of dihedral angles between the methyl carbon atoms of these O- and N-acetyl groups (that we have arbitrarily named *n* and *p*) indicates that the majority of the population of O25a conformers favor this staggered arrangement (Figure 31).



**Figure 31:** An analysis of an arbitrary pair of dihedral angles, *n* and *p*, between the methyl carbon atoms of the O- and N- acetyl groups of the O25a trimer. (a) Probability density plot of *n* and *p*. (b) Representative O25a conformers showing the selected angles. Note the staggered arrangement of the O- and N- acetyl groups.

Hydrogen bonding between the C (4) hydroxyl groups of N-acetyl-glucosamine monosaccharides and the O (5) ring atoms of adjacent N-acetyl-fucosamine monosaccharides appear to stabilize the staggered conformation (Appendix I). Hydrogen bonds between the primary hydroxyl groups of the  $\beta$ -D-glucopyranose side chains and the carbonyl oxygen atoms of the main-chain N-acetyl-D-glucosamine monosaccharides that stabilise the compressed O25b conformation are absent. We postulate that the aG13aFNAc bond is not flexible enough to accommodate these interactions.

Finally, the preferred dihedral values for the central O25a RU are in agreement with those necessary for a staggered, linear conformation of the O25a oligosaccharide – refer to Table 4.

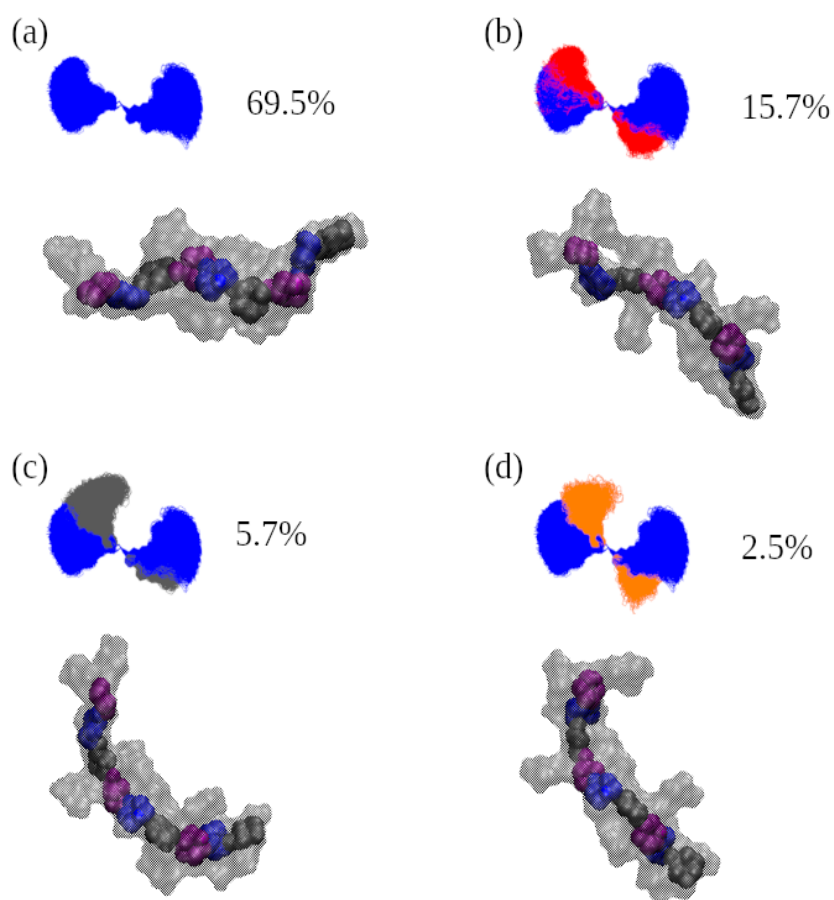
Our comparative analysis of the preferred dihedral values of the O25a and O25b trimers provides a rationale for the differences in their predicted levels of extension: First, O- acetylation of main-chain  $\alpha$ -L-rhamnopyranose monosaccharides enhances the flexibility of the aG13aRAc bond in O25b. This facilitates hydrogen bonding between the primary hydroxyl groups of  $\beta$ -D-glucopyranose side-chains and the carbonyl oxygen atoms of main-chain N-acetyl-D-glucosamine monosaccharides. These interactions stabilize a large population of compressed, helical O25b conformers with end-to-end distances of around 13.7 Å. Antagonistic hydrogen bonds stabilize an alternative population of more extended, loosely helical O25b conformers with end-to-end distances of about 21 Å. Second, O- and N-acetylation in O25a restrict rotation about its aG13aFNAc and aFNAc13bGNAc linkages. These bonds are not flexible enough to accommodate the side-chain to main-chain interactions predicted in O25b. Instead, the O25a trimer adopts a more rigid, linear conformation with the O- and N-acetyl groups of adjacent RUs in a staggered arrangement about the vertical axis of the molecule.

### 7.2.3 RMSD cluster analysis of O25a and O25b trimers

Root Mean Square Deviation (RMSD) analysis provides a measure of the overall variation of a population of conformers from one another and is a useful metric for grouping related conformations into clusters. RMSD clustering of 3 RU oligosaccharide extensions of O25a and O25b is presented here as an extension of the discussion covered in Sections 7.2.1 and 7.2.2 and to provide an additional view of the conformations predicted for the trimers with their relative weightings.

Each oligosaccharide extensions is aligned to the central monosaccharide of their central RUs (aFNAc in O25a and aRAc in O25b). These units were selected as they are conveniently located at the center of each trimer and we have established that their glycosidic bonds are key determinants of the overall conformations of these molecules (Refer specifically to the highlighted dihedral angles presented in Table 4). We have disregarded the first 100 ns of simulation time for equilibration and focus our analysis on the last 100 ns sampling conformers every 2.5 ps.

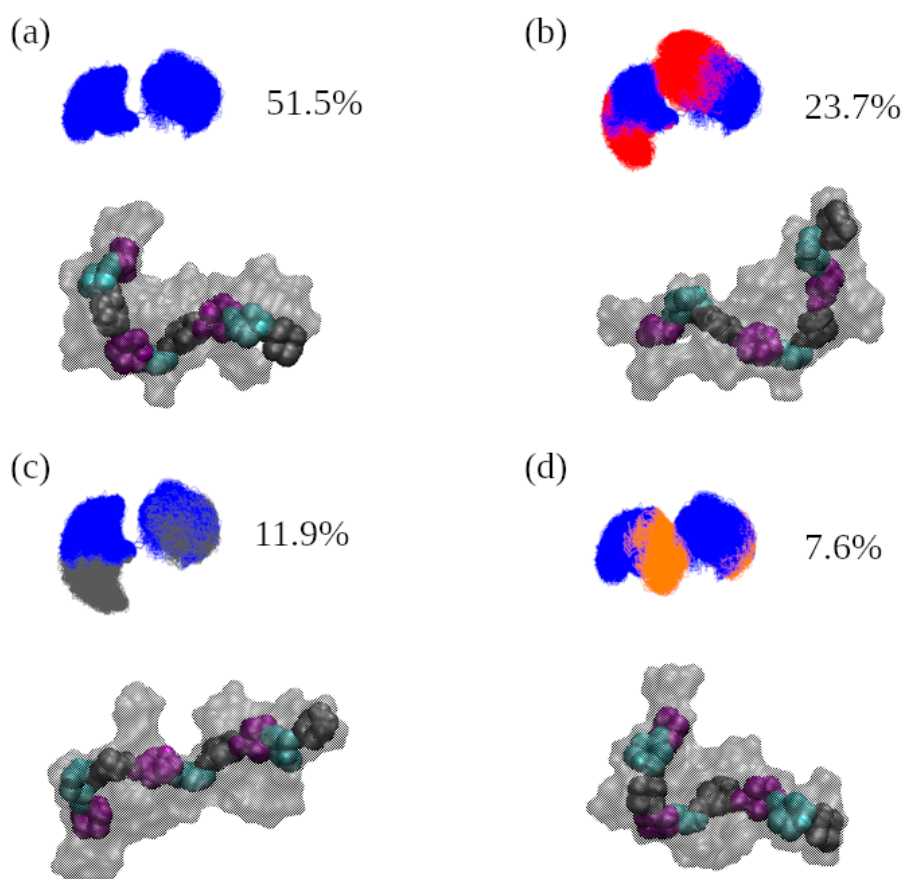
Clusters, weightings and representative conformers of the O25a trimer are presented in Figure 32:



*Figure 32: RMSD clustering of a 3 RU O25a oligosaccharide extension aligned around the central  $\alpha$ LFuc monosaccharide. 4 clusters with a cutoff value of 4 were selected grouping 93.4% of conformers. All views are along the z-axis with the non-reducing end on the left. Clustered is based on the arrangement of backbone monosaccharides as highlighted. (a) Shows the arrangement, weighting and a representative conformer for the largest cluster. (b), (c), (d) Show the arrangements, weightings and conformers of the other 3 clusters in relation to (a).  $\alpha$ -D-Glc in purple,  $\beta$ -D-GlcNAc in grey,  $\alpha$ -L-FucNAc in blue.*

Although the selection of 4 clusters is somewhat arbitrary, the conformers presented in Figure 32 are in agreement with the level of extension and dihedral angle discussions presented in Sections 7.2.1 and 7.2.2. An overwhelming majority of conformers adopt an extended, somewhat curved arrangement that varies around the central fucose monosaccharide in relation to the  $\psi$  dihedral angle populations presented in Figure 27. Smaller clusters of conformers that exhibit somewhat more curved arrangements are evident but with significantly lower weightings. These O25a conformers are, as expected, notably different to those presented for O25b in Figure 33:





**Figure 33:** RMSD clustering of a 3 RU O25b oligosaccharide extension aligned around the central *aLRha* monosaccharide. 4 clusters with a cutoff value of 4.5 were selected grouping 94.7% of conformers. All views are along the z-axis with the non-reducing end on the left. Backbone monosaccharides were selected for clustering and are highlighted. (a) Shows the arrangement, weighting and a representative conformer for the largest cluster. (b), (c), (d) Show the arrangements, weightings and conformers of the other 3 clusters in relation to (a).  *$\alpha$ -D-Glc* in purple,  *$\beta$ -D-GlcNAc* in grey,  *$\alpha$ -L-RhaAc* in cyan.

The tight helical conformation postulated for O25b in Sections 7.2.1 and 7.2.2 is clearly visible in the conformers presented in Figure 33: an overwhelming majority exhibit these right-handed helical arrangements. A much smaller population (11.9 %) appear to present the extended conformation (stabilised by antagonistic hydrogen bonds) presented in Section 7.2.2. Although the 2- N-acetyl and O-acetyl side groups are not explicitly highlighted here, N-acetyl groups (at C2 of  *$\beta$ -D-Glc* units) are rotated toward the center of the helix in the majority of conformers and acetyl groups (at C 2 of  *$\alpha$ -L-Rha* units) are rotated towards the outside of the molecule. This arrangement presents a markedly different epitope to the O25a antigen.

#### 7.2.4 Concluding remarks

Our results indicate that the preferred conformations predicted for trimers of O25a and O25b differ significantly in both their levels of extensions and their geometric shapes in solution.

Conformational differences between bacterial O-antigens are known to effect their cross reactivity with related antibodies.<sup>22</sup> The geometric differences between the O25a and O25b trimers noted here provide a rationale for the existence of monoclonal antibodies that react with O25b but largely lack cross reactivity to O25a.<sup>9-11</sup>

Furthermore, the N-acetyl groups of O25b are predicted to be involved in intramolecular hydrogen bonds and are rotated toward the reducing to non-reducing end axis of a collapsed, helical molecule in the dominant antigen conformation. This may serve to shield N-acetyl and expose O-acetyl groups on the surface of O25b.

Our predicted conformation for the 3 RU O25a oligosaccharide exposes pairs of O- and N-acetyl groups in close proximity to one another on the antigen surface. This arrangement is markedly different to that predicted for O25b - even when O25b is in the extended conformation that we have described. The accessibility of O-acetyl groups has been shown to affect antigen-antibody binding<sup>22</sup> and thus presents another possible explanation for the observed clinical differences in monoclonal antibody cross reactivity between O25a and O25b.



## Chapter 8: Conclusions and future work

This study presented an initial computational investigation of the conformations of the *E. coli* O25a and O25b polysaccharide antigens as well as a case study comparing the CHARMM36 v1.126 and GLYCAM06j carbohydrate force fields.

Despite some differences, our CHARMM36 and GLYCAM06 simulations were largely in agreement with regards to the conformation of a 3 RU oligosaccharide extension of the classical *E. coli* O25 antigen without O- or N-acetyl groups: both force fields predicted a fairly linear, extended conformation of this molecule in solution.

CHARMM36 and GLYCAM06 predicted markedly different conformations for a 3 RU oligosaccharide extension of the O25b antigen without O- or N-acetyl groups: CHARMM36 indicated an extended, linear molecule whereas GLYCAM06 predicted the formation of a collapsed, globular conformation that was not reversible over the simulation duration. The collapsed O25b conformation predicted by GLYCAM06 was associated with a tight, high-energy, hairpin-like bend at a main-chain  $\alpha$ -L-rhamnopyranose-(1->3)- $\beta$ -D-glucopyranose linkage and a highly constrained  $\beta$ -D-glucopyranose side chain in an anti- $\Psi$ , anti- $\omega$  orientation. These features were absent from the CHARMM36 simulation data. Other studies have reported the collapse of GLYCAM06 oligosaccharide simulations into similar “metastable” conformations and, whilst clear empirical evidence is required to ascertain which conformation reflects reality, the globular O25b conformation predicted by GLYCAM06 may not be reliable.

Simulation with the CHARMM36 force field predicted that O- and N-acetylation have a small extending effect on a 3 RU oligosaccharide extension of the classical O25 *E. coli* antigen and a much more pronounced compressive effect on a 3 RU O25b oligosaccharide extension. We presented evidence that 2 O-acetylation of  $\alpha$ -L-rhamnopyranose units in O25b facilitates rotation about main-chain  $\alpha$ -D-glucopyranose-(1->3)-2O-acetyl-rhamnopyranose bonds into orientations that are constrained in the corresponding  $\alpha$ -D-glucopyranose-(1->3)-N-acetyl-fucosamine bonds of O25. This allows the O25b trimer to arrange into a compressed helical conformation stabilized by hydrogen bonds between centrally orientated N-acetyl groups and  $\beta$ -D-glucopyranose side chains. The compressed helix of O25b may shield N-acetyl groups around the inside of the helix and expose O-acetyl groups around the outside. An alternative population of conformers of the O25b

trimer with a loosely helical conformation stabilized by antagonistic hydrogen bonds was also described. CHARMM36 simulation predicted that a classical O25 trimer has a fairly linear, extended conformation in solution. This conformation was found to involve a staggered arrangement of pairs of N-acetyl and O-acetyl groups facing away from the vertical axis of the molecule.

Overall, CHARMM36 simulation predicted large conformational differences between trimers of the classical O25 *E. coli* antigen and the more recently characterized O25b antigen. These include geometric differences in the general shapes of the molecules and differences in the orientations and predicted levels of exposure of O- and N-acetyl groups. We postulate that these large conformational differences provide a rationale for the existence of monoclonal antibodies that are largely selective for the O25b antigen.

Longer molecular dynamics simulations, and simulations of longer oligosaccharide extensions, may be useful in providing an enhanced understanding of the structure of the O25 and O25b antigens *in vivo*. In this investigation we have assumed that all pyranoses exist in “chair conformations” in solution. A more detailed investigation into the effects of ring-pucker on conformation may be of future interest. Membrane studies of the O25b antigen that make use of new tools such as the CHARMM-GUI LPS Builder would be of interest considering the predicted conformational differences between the classical O25 and O25b antigens. Lastly, most pathogenic ST131 *E. coli* strains express the O25b polysaccharide but some  $\beta$ -lactam resistant bacteria that express the O16 antigen have also been reported. A similar analysis of the O16 polysaccharide, therefore, presents an opportunity for additional work on ST131:H4 bacteria.

## Chapter 9: References

- (1) . Stahle J., Widmalm G., Kuttel M. M. CarbBuilder: Software for Building Molecular Models of Complex Oligo- and Polysaccharide Structures. *Journal of Computational Chemistry* **2016**, 37 (22), 2098–2105. <https://doi.org/10.1002/jcc.24428>.
- (2) Kuttel M., Gordon M., Ravenscroft N. Comparative Simulation of Pneumococcal Serogroup 19 Polysaccharide Repeating Units with Two Carbohydrate Force Fields. *Carbohydrate Research* **2014**.
- (3) ...Kuttel M., Timol Z., Ravenscroft N. Cross-Protection in Neisseria Meningitidis Serogroups Y and W Polysaccharides: A Comparative Conformational Analysis. *Carbohydrate Research* **2017**, 446–447, 40–47. <https://doi.org/10.1016/j.carres.2017.05.004>.
- (4) .....Kuttel M., Cescutti P., Distefano M., and Rizzo R. Fluorescence and NMR Spectroscopy Together with Molecular Simulations Reveal Amphiphilic Characteristics of a Burkholderia Biofilm Exopolysaccharide. *JBC* **2017**, 292 (26), 11034–11042. <https://doi.org/10.1074/jbc.M117.785048>.
- (5) ....Solomons, T. W. G.; Fryhle, C. B. *Organic Chemistry*, 9 edition.; Wiley: Hoboken, NJ, 2007.
- (6) . Woods, R. J. Computational Carbohydrate Chemistry: What Theoretical Methods Can Tell Us. *Glycoconj J* **1998**, 15 (3), 209–216.
- (7) Aslam, B.; Wang, W.; Arshad, M. I.; Khurshid, M.; Muzammil, S.; Rasool, M. H.; Nisar, M. A.; Alvi, R. F.; Aslam, M. A.; Qamar, M. U.; Salamat, M. K. F.; Baloch, Z. Antibiotic Resistance: A Rundown of a Global Crisis. *Infect Drug Resist* **2018**, 11, 1645–1658. <https://doi.org/10.2147/IDR.S173867>.
- (8) .....Vliegthart, J. F. G. Carbohydrate Based Vaccines. *FEBS Lett.* **2006**, 580 (12), 2945–2950. <https://doi.org/10.1016/j.febslet.2006.03.053>.
- (9) .....Szijártó V., Lukasiewicz J., Gozdiewicz T. K., Magyarics Z., Nagy E., Nagy G. Diagnostic Potential of Monoclonal Antibodies Specific to the Unique O-Antigen of Multidrug-Resistant Epidemic Escherichia Coli Clone ST131-O25b:H4. *CVI* **2014**, 21 (7), 930–939. <https://doi.org/10.1128/CVI.00685-13>.
- (10) .....Szijártó V., Guachalla L. M., Visram Z. C., Hartl K., Varga C., Mirkina I., Zmajkovic J., Badarau A.; Gerhild Zauner, Clara Pleban, Zoltán Magyarics, Eszter Nagy, Gábor Nagy. Bactericidal Monoclonal Antibodies Specific to the Lipopolysaccharide O Antigen from Multidrug-Resistant Escherichia Coli Clone ST131- O25b:H4 Elicit Protection in Mice. *AAC* **2015**, 59 (6), 3109–3116. <https://doi.org/10.1128/AAC.04494-14>.

- (11) Guachalla L. M., Hartl K., Varga C., Stulik L., Mirkina I.,; Stefan Malafa, Eszter Nagy, Gábor Nagy, Valéria Szijártó. Multiple Modes of Action of a Monoclonal Antibody against Multidrug-Resistant Escherichia Coli Sequence Type 131-H30. *AAC* **2017**, 61 (11), 1–11.  
<https://doi.org/10.1128/AAC.01428-17>.
- (12) .....Cramer C. J. *Essentials of Computational Chemistry: Theories and Models, 2nd Edition*, Second edition.; John Wiley & Sons, 2004.
- (13) Ciccotti G., Ferrario M. and Schuette C. *Molecular Dynamics Simulation*, First edition.; MDPI, 2014.
- (14) . .Vila, J.; Sáez-López, E.; Johnson, J. R.; Römling, U.; Dobrindt, U.; Cantón, R.; Giske, C. G.; Naas, T.; Carattoli, A.; Martínez-Medina, M.; Bosch, J.; Retamar, P.; Rodríguez-Baño, J.; Baquero, F.; Soto, S. M. Escherichia Coli: An Old Friend with New Tidings. *FEMS Microbiol Rev* **2016**, 40 (4), 437–463. <https://doi.org/10.1093/femsre/fuw005>.
- (15) .....Blount, Z. D. The Unexhausted Potential of E. Coli. *eLife* 4. <https://doi.org/10.7554/eLife.05826>.
- (16) Stenutz R., Weintraub A. & Widmalm G. The Structures of Escherichia Coli O-Polysaccharide Antigens. *FEMS* **2006**, 30, 382–403. <https://doi.org/10.1111/j.1574-6976.2006.00016.x>.
- (17) ....Nicolas-Chanoine, M.-H.; Bertrand, X.; Madec, J.-Y. Escherichia Coli ST131, an Intriguing Clonal Group. *Clin. Microbiol. Rev.* **2014**, 27 (3), 543–574. <https://doi.org/10.1128/CMR.00125-13>.
- (18) .....Cagnacci S., Gualco L., Debbia E., Schito G. C., and Marchese A. European Emergence of Ciprofloxacin-Resistant Escherichia Coli Clonal Groups O25:H4-ST 131 and O15:K52:H1 Causing Community-Acquired Uncomplicated Cystitis. *Journal of Clinical Microbiology* **2008**, 46 (8), 2605–2612. <https://doi.org/10.1128/JCM.00640-08>.
- (19) .....Rogers BA, Sidjabat HE, Paterson DL. Escherichia Coli O25b-ST131: A Pandemic, Multiresistant, Community-Associated Strain. *J Antimicrob Chemother* **2011**, 66 (1), 1–14. <https://doi.org/10.1093/jac/dkq415>.
- (20) ...Haji-Ghassemi, O.; Blackler, R. J.; Martin Young, N.; Evans, S. V. Antibody Recognition of Carbohydrate Epitopes. *Glycobiology* **2015**, 25 (9), 920–952. <https://doi.org/10.1093/glycob/cwv037>.
- (21) .....Kuttel M. M., Jackson G. E., Mafata M., Ravenscroft N. Capsular Polysaccharide Conformations in Pneumococcal Serotypes 19F and 19A. *Carbohydrate Research* **2015**, 406, 27–33. <https://doi.org/10.1016/j.carres.2014.12.013>.

- (22) ....Kuttel, M. M.; Ravenscroft, N. The Role of Molecular Modeling in Predicting Carbohydrate Antigen Conformation and Understanding Vaccine Immunogenicity. In *Carbohydrate-Based Vaccines: From Concept to Clinic*; ACS Symposium Series; American Chemical Society, 2018; Vol. 1290, pp 139–173. <https://doi.org/10.1021/bk-2018-1290.ch007>.
- (23) .....Zhu X., Lopes P. E. M., MacKerell A. D. Jr. Recent Developments and Applications of the CHARMM Force Fields. *Wiley Interdiscip Rev Comput Mol Sci* **2011**, 2 (1), 167–185. <https://doi.org/10.1002/wcms.74>.
- (24) Hatcher, E., Guvench, O., and MacKerell, Jr., A.D. CHARMM Additive All-Atom Force Field for Acyclic Polyalcohols, Acyclic Carbohydrates and Inositol. *Journal of Computational Chemistry* **2009**, 5 (5), 1315–1327. <https://doi.org/10.1021/ct9000608>.
- (25) ....Hatcher, E.; Guvench, O.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Aldopentofuranoses, Methyl-Aldopentofuranosides, and Fructofuranose. *J. Phys. Chem. B* **2009**, 113 (37), 12466–12476. <https://doi.org/10.1021/jp905496e>.
- (26) . Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Carbohydrate Derivatives and Its Utility in Polysaccharide and Carbohydrate–Protein Modeling. *J. Chem. Theory Comput.* **2011**, 7 (10), 3162–3180. <https://doi.org/10.1021/ct200328p>.
- (27) . Raman, E. P.; Guvench, O.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages in Carbohydrates Involving Furanoses. *J. Phys. Chem. B* **2010**, 114 (40), 12981–12994. <https://doi.org/10.1021/jp105758h>.
- (28) .....Mallajosyula, S. S.; Guvench, O.; Hatcher, E.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Phosphate and Sulfate Linked to Carbohydrates. *J Chem Theory Comput* **2012**, 8 (2), 759–776. <https://doi.org/10.1021/ct200792v>.
- (29) .....Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J Comput Chem* **2008**, 29 (4), 622–655. <https://doi.org/10.1002/jcc.20820>.
- (30) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99 (20), 12562–12566. <https://doi.org/10.1073/pnas.202427399>.
- (31) .....Barducci, A., Massimiliano Bonomi, M. & Parrinello, M. Metadynamics WIREs Computational Molecular Science Volume 1.
- (32) . Salton, M. R. J.; Kim, K.-S. Structure. In *Medical Microbiology*; Baron, S., Ed.; University of Texas Medical Branch at Galveston: Galveston (TX), 1996.

- (33) .....Bitzer, R. S.; Barbosa, A. G. H.; da Silva, C. O.; Nascimento, M. A. C. On the Generalized Valence Bond Description of the Anomeric and Exo-Anomeric Effects: An Ab Initio Conformational Study of 2-Methoxytetrahydropyran. *Carbohydr. Res.* **2005**, *340* (13), 2171–2184. <https://doi.org/10.1016/j.carres.2005.07.001>.
- (34) ..Salisbury, A. M.; Deline, A. L.; Lexa, K. W.; Shields, G. C.; Kirschner, K. N. Ramachandran-Type Plots for Glycosidic Linkages: Examples from Molecular Dynamic Simulations Using the Glycam06 Force Field. *Journal of Computational Chemistry* **2009**, *30* (6), 910–921. <https://doi.org/10.1002/jcc.21099>.
- (35) .....Boltzmann Distribution. *Wikipedia*; 2019.
- (36) .....Guvench O., Hatcher E., Venable R. M., Pastor R. W., MacKerell A. D. Jr. Additive Empirical CHARMM Force Field for Glycosyl Linked Hexopyranoses. *J. Chem. Theory Comput.* **2009**, *5* (9), 2353–2370. <https://doi.org/10.1021/ct900242e>.
- (37) ..Woods Research Group. *GLYCAM Web*; Complex Carbohydrate Research Center, University of Georgia, Athens, GA., 2005.
- (38) .....French A. D. Chapter 21: Energy Maps for Glycosidic Linkage Conformations. *Glycoinformatics, Methods in Molecular Biology* **2015**, *1273*, 333–357. [https://doi.org/10.1007/978-1-4939-2343-4\\_21](https://doi.org/10.1007/978-1-4939-2343-4_21).
- (39) Kong K., Schnepfer L., and Mathee K. Beta-Lactam Antibiotics: From Antibiosis to Resistance and Bacteriology. *APMIS* **201AD**, *118* (1), 1–36. <https://doi.org/10.1111/j.1600-0463.2009.02563.x>.
- (40) .....Giedraitienė, A.; Vitkauskienė, A.; Naginienė, R.; Pavilonis, A. Antibiotic Resistance Mechanisms of Clinically Important Bacteria. *Medicina* **2011**, *47* (3), 19. <https://doi.org/10.3390/medicina47030019>.
- (41) .....Katzenellenbogen, E.; Kocharova, N. A.; Zatonsky, G. V.; Shashkov, A. S.; Bogulska, M.; Knirel, Y. A. Structures of the Biological Repeating Units in the O-Chain Polysaccharides of *Hafnia Alvei* Strains Having a Typical Lipopolysaccharide Outer Core Region. *FEMS Immunol Med Microbiol* **2005**, *45* (2), 269–278. <https://doi.org/10.1016/j.femsim.2005.05.003>.
- (42) ...Li, C.; Diprimio, N.; Bowles, D. E.; Hirsch, M. L.; Monahan, P. E.; Asokan, A.; Rabinowitz, J.; Agbandje-McKenna, M.; Samulski, R. J. Single Amino Acid Modification of Adeno-Associated Virus Capsid Changes Transduction and Humoral Immune Profiles. *Journal of Virology* **2012**, *86* (15), 7752–7759. <https://doi.org/10.1128/JVI.00675-12>.

- (43) .....Weintraub A. Immunology of Bacterial Polysaccharide Antigens. *Carbohydrate Research* **2003**, 338, 2539–2547. <https://doi.org/10.1016/j.carres.2003.07.008>.
- (44) .....Tanis T.C., Duncan D., MacCannell R.†. Molecular Strain Typing and Characterisation of Toxigenic *Clostridium Difficile*. *Methods in Microbiology* **2015**, 42, 329–357.
- (45) Nicolas-Chanoine, M.-H.; Blanco, J.; Leflon-Guibout, V.; Demarty, R.; Alonso, M. P.; Caniça, M. M.; Park, Y.-J.; Lavigne, J.-P.; Pitout, J.; Johnson, J. R. Intercontinental Emergence of *Escherichia Coli* Clone O25:H4-ST131 Producing CTX-M-15. *J. Antimicrob. Chemother.* **2008**, 61 (2), 273–281. <https://doi.org/10.1093/jac/dkm464>.
- (46) .....Peirano, G.; Bradford, P. A.; Kazmierczak, K. M.; Badal, R. E.; Hackel, M.; Hoban, D. J.; Pitout, J. D. D. Global Incidence of Carbapenemase-Producing *Escherichia Coli* ST131. *Emerg Infect Dis* **2014**, 20 (11), 1928–1931. <https://doi.org/10.3201/eid2011.141388>.
- (47) .....Dahbi, G.; Mora, A.; López, C.; Alonso, M. P.; Mamani, R.; Marzoa, J.; Coira, A.; García-Garrote, F.; Pita, J. M.; Velasco, D.; Herrera, A.; Viso, S.; Blanco, J. E.; Blanco, M.; Blanco, J. Emergence of New Variants of ST131 Clonal Group among Extraintestinal Pathogenic *Escherichia Coli* Producing Extended-Spectrum  $\beta$ -Lactamases. *Int. J. Antimicrob. Agents* **2013**, 42 (4), 347–351. <https://doi.org/10.1016/j.ijantimicag.2013.06.017>.
- (48) .....Mathers, A. J.; Peirano, G.; Pitout, J. D. D. Chapter Four - *Escherichia Coli* ST131: The Quintessential Example of an International Multiresistant High-Risk Clone. In *Advances in Applied Microbiology*; Sariaslani, S., Gadd, G. M., Eds.; Academic Press, 2015; Vol. 90, pp 109–154. <https://doi.org/10.1016/bs.aambs.2014.09.002>.
- (49) . Kenne, L.; Lindberg, B.; Madden, J. K.; Lindberg, A. A.; Gemski, P. Structural Studies of the *Escherichia Coli* O-Antigen 25. *Carbohydrate Research* **1983**, 122 (2), 249–256. [https://doi.org/10.1016/0008-6215\(83\)88336-0](https://doi.org/10.1016/0008-6215(83)88336-0).
- (50) Rojas-Macias, M. A.; Ståhle, J.; Lütteke, T.; Widmalm, G. Development of the ECODAB into a Relational Database for *Escherichia Coli* O-Antigens and Other Bacterial Polysaccharides. *Glycobiology* **2015**, 25 (3), 341–347. <https://doi.org/10.1093/glycob/cwu116>.
- (51) . Perez, S., Sarkar, A., Rivet, A., Drouillard, S., Breton, C. & Imberty, A. Glyco3D: A Suite of Interlinked Databases of 3D Structures of Complex Carbohydrates, Lectins, Antibodies, and Glycosyltransferases. *A Practical Guide to Using Glycomics Databases* **2017**, 133–161.
- (52) Böhm M, Bohne-Lang A, Frank M, Loss A, Rojas-Macias MA, Lütteke T. Glycosciences.DB: An Annotated Data Collection Linking Glycomics and Proteomics Data (2018 Update). *Nucleic Acids Res.* **2019**, 47 (D1), D1195–D1201. <https://doi.org/10.1093/nar/gky994>.



- (53) . Vilchez, S.; Lundborg, M.; Urbina, F.; Weintraub, A.; Widmalm, G. Structural Studies of the O-Antigenic Polysaccharides from the Enteropathogenic Escherichia Coli Strain 94/D4 and the International Type Strain Escherichia Coli O82. *Carbohydrate Research* **2009**, *344* (18), 2528–2532. <https://doi.org/10.1016/j.carres.2009.09.033>.
- (54) .....Svensson, M. V.; Weintraub, A.; Widmalm, G. Structural Elucidation of the O-Antigenic Polysaccharide from Escherichia Coli O175. *Carbohydrate Research* **2011**, *346* (3), 449–453. <https://doi.org/10.1016/j.carres.2010.12.005>.
- (55) .....Perepelov, A. V.; Shashkov, A. S.; Guo, X.; Filatov, A. V.; Weintraub, A.; Widmalm, G.; Knirel, Y. A. Structure and Genetics of the O-Antigen of Escherichia Coli O169 Related to the O-Antigen of Shigella Boydii Type 6. *Carbohydrate Research* **2015**, *414*, 46–50. <https://doi.org/10.1016/j.carres.2015.05.016>.
- (56) .....Masoud, H.; Perry, M. B. Structural Characterization of the O-Antigenic Polysaccharide of Escherichia Coli Serotype O17 Lipopolysaccharide. *Biochem. Cell Biol.* **1996**, *74* (2), 241–248. <https://doi.org/10.1139/o96-025>.
- (57) . Manca, M. C.; Weintraub, A.; Widmalm, G. Structural Studies of the Escherichia Coli O26 O-Antigen Polysaccharide. *Carbohydrate Research* **1996**, *281* (1), 155–160. [https://doi.org/10.1016/0008-6215\(95\)00333-9](https://doi.org/10.1016/0008-6215(95)00333-9).
- (58) ...Fontana, C.; Weintraub, A.; Widmalm, G. Structural Studies and Biosynthetic Aspects of the O-Antigen Polysaccharide from Escherichia Coli O42. *Carbohydrate Research* **2015**, *403*, 174–181. <https://doi.org/10.1016/j.carres.2014.05.003>.
- (59) . Fontana, C.; Lundborg, M.; Weintraub, A.; Widmalm, G. Structural Studies and Biosynthetic Aspects of the O-Antigen Polysaccharide from Escherichia Coli O174. *Carbohydrate Research* **2012**, *354*, 102–105. <https://doi.org/10.1016/j.carres.2012.02.020>.
- (60) .....Ali, T.; Weintraub, A.; Widmalm, G. Structural Determination of the O-Antigenic Polysaccharide from Escherichia Coli O166. *Carbohydrate Research* **2007**, *342* (2), 274–278. <https://doi.org/10.1016/j.carres.2006.11.023>.
- (61) Rosen, J.; Robobi, A.; Nyholm, P.-G. The Conformations of the O-Specific Polysaccharides of Shigella Dysenteriae Type 4 and Escherichia Coli O159 Studied with Molecular Mechanics (MM3) Filtered Systematic Search. *Carbohydrate Research* **2004**, *339* (5), 961–966. <https://doi.org/10.1016/j.carres.2003.11.018>.
- (62) .....Ghosh, T.; Kar, R. K.; Bhunia, A.; Misra, A. K. Synthesis of the Pentasaccharide Repeating Unit of the O-Antigen of Escherichia Coli O175 Using One-Pot Glycosylations and Its

Conformational Analysis. *Tetrahedron* **2014**, 70 (48), 9262–9267.

<https://doi.org/10.1016/j.tet.2014.09.080>.

(63) . .Wu, E. L.; Fleming, P. J.; Yeom, M. S.; Widmalm, G.; Klauda, J. B.; Fleming, K. G.; Im, W. E. Coli Outer Membrane and Interactions with OmpLA. *Biophysical Journal* **2014**, 106 (11), 2493–2502. <https://doi.org/10.1016/j.bpj.2014.04.024>.

(64) Blasco, P.; Patel, D. S.; Engström, O.; Im, W.; Widmalm, G. Conformational Dynamics of the Lipopolysaccharide from Escherichia Coli O91 Revealed by Nuclear Magnetic Resonance Spectroscopy and Molecular Simulations. *Biochemistry* **2017**, 56 (29), 3826–3839. <https://doi.org/10.1021/acs.biochem.7b00106>.

(65) ...Lee, J.; Patel, D. S.; Stähle, J.; Park, S.-J.; Kern, N. R.; Kim, S.; Lee, J.; Cheng, X.; Valvano, M. A.; Holst, O.; Knirel, Y. A.; Qi, Y.; Jo, S.; Klauda, J. B.; Widmalm, G.; Im, W. CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *J. Chem. Theory Comput.* **2019**, 15 (1), 775–786. <https://doi.org/10.1021/acs.jctc.8b01066>.

(66) Laio A. and Gervasio F. L. Metadynamics: A Method to Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry and Material Science. *Rep. Prog. Phys.* 71 (12). <https://doi.org/10.1088/0034-4885/71/12/126601>.

(67) ....Phillips J. C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R. D., Kale L., Schulten K. Scalable Molecular Dynamics with NAMD. *Journal of Computational Chemistry* **2005**, 26, 1781–1802.

(68) .....Humphrey W, Dalke A, Schulten K. VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, 14, 33–38.

(69) .....AMBER; University of California, San Francisco.

(70) .....Williams, T.; Kelley, C.; many others. *Gnuplot 4.6: An Interactive Plotting Program*; 2013.

(71) .....Unknown. ImageMagick website <http://www.imagemagick.org/>.

(72) .....Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Jarrod Millman, K.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C.; Polat, İ.; Feng, Y.; Moore, E. W.; Vand erPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Contributors, S. 1. 0. SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python. *arXiv e-prints* **2019**, arXiv:1907.10121.

- (73) ...Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **2007**, 9 (3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
- (74) .....Kang, Y.; Barbirz, S.; Lipowsky, R.; Santer, M. Conformational Diversity of O-Antigen Polysaccharides of the Gram-Negative Bacterium *Shigella Flexneri* Serotype Y. *J. Phys. Chem. B* **2014**, 118 (9), 2523–2534. <https://doi.org/10.1021/jp4111713>.
- (75) .....Imberty, A. Oligosaccharide Structures: Theory versus Experiment. *Current Opinion in Structural Biology* **1997**, 7 (5), 617–623. [https://doi.org/10.1016/S0959-440X\(97\)80069-3](https://doi.org/10.1016/S0959-440X(97)80069-3).
- (76) Zierke, M.; Smieško, M.; Rabbani, S.; Aeschbacher, T.; Cutting, B.; Allain, F. H.-T.; Schubert, M.; Ernst, B. Stabilization of Branched Oligosaccharides: Lewisx Benefits from a Nonconventional C–H···O Hydrogen Bond. *J. Am. Chem. Soc.* **2013**, 135 (36), 13464–13472. <https://doi.org/10.1021/ja4054702>.
- (77) ...Battistel M.D., Azurmendi H.F., Frank M., Freedberg D.I. Uncovering Nonconventional and Conventional Hydrogen Bonds in Oligosaccharides through NMR Experiments and Molecular Modeling: Application to Sialyl Lewis-X. *J Am Chem Soc.* **2015**, 137 (42), 13444–13447. <https://doi.org/doi:10.1021/jacs.5b03824>. Epub 2015 Oct 15.
- (78) . Horton, D. Anomeric Effect in Sugars. In *Encyclopedia of Biophysics*; Roberts, G. C. K., Ed.; Springer: Berlin, Heidelberg, 2013; pp 93–94. [https://doi.org/10.1007/978-3-642-16712-6\\_69](https://doi.org/10.1007/978-3-642-16712-6_69).

## Appendix A: Analysis of PMF contour plots of disaccharide fragments

**Table A1:** Dihedral values calculated from Metadynamics simulation of O25a' disaccharides in vacuum. Actual starting structure dihedrals are indicated in brackets if notably different.

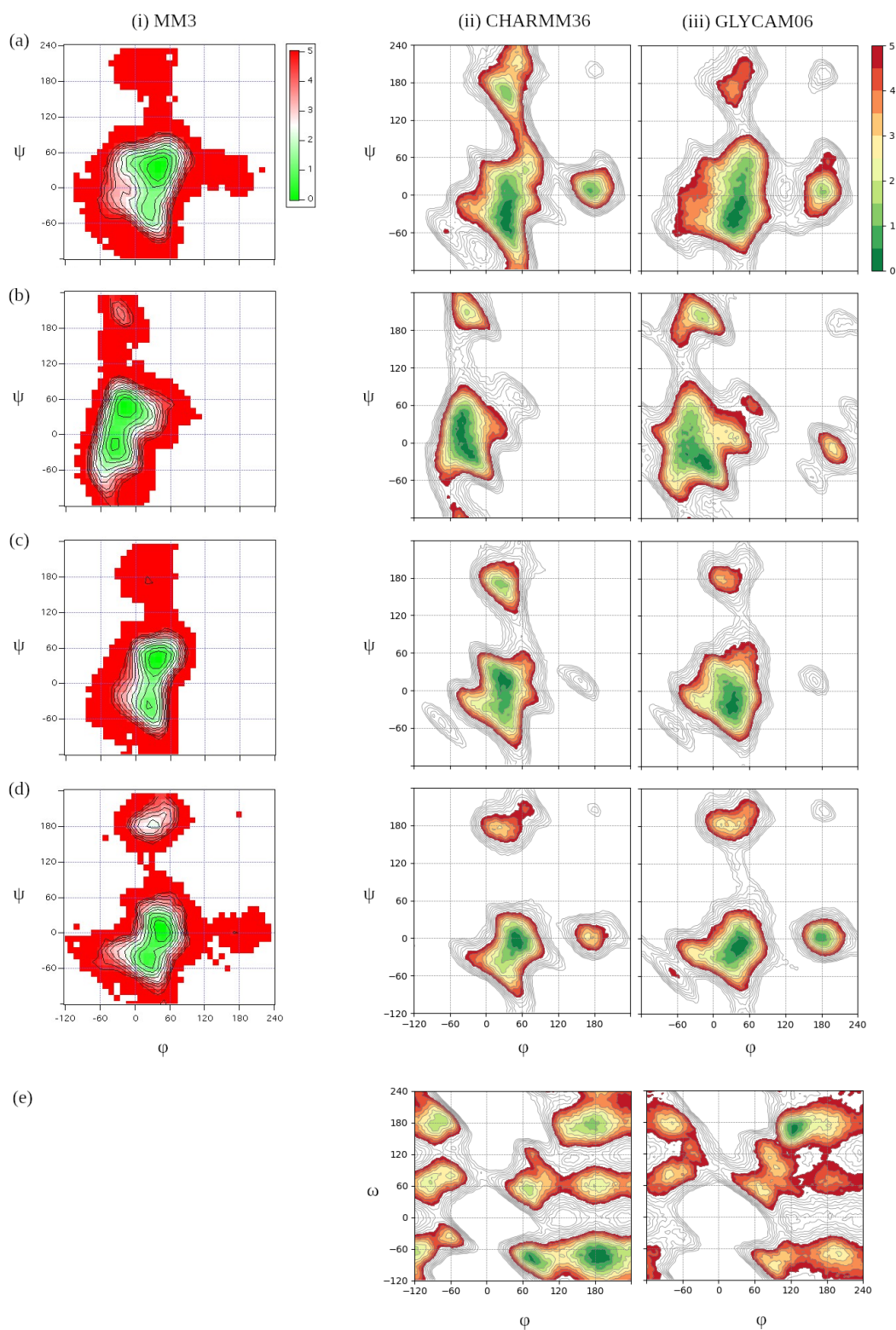
Glycosidic Bonds	CHARMM36 minima			GLYCAM06 minima		
	$\varphi / ^\circ$	$\Psi / ^\circ$	$\omega / ^\circ$	$\varphi / ^\circ$	$\Psi / ^\circ$	$\omega / ^\circ$
$\alpha$ -L-Rhap-(1->3)- $\alpha$ -D-Glcp	36 (34)	-36 (166)	-	39 (43)	-29 (155)	-
$\beta$ -D-Glcp-(1->6)- $\alpha$ -D-Glcp	(0)	-176	-78 (-60)	29 (42)	126 (43)	166 (171)
$\beta$ -D-Glcp-(1->4)- $\alpha$ -D-Glcp	54 (47)	-6 (3)	-	49 (39)	-11 (-19)	-
$\alpha$ -D-Glcp-(1->3)- $\alpha$ -L-Fucp	-31	-14	-	-16 (-8)	-26 (-36)	-
$\alpha$ -L-Fucp-(1->3)- $\beta$ -D-Glcp	29	14	-	29 (38)	-24 (-40)	-

**Table A2:** Dihedral values calculated from Metadynamics simulation of O25b' disaccharides in vacuum. Actual starting structure dihedrals are indicated in brackets if notably different.

Glycosidic Bonds	CHARMM36 minima			GLYCAM06 minima		
	$\varphi / ^\circ$	$\Psi / ^\circ$	$\omega / ^\circ$	$\varphi / ^\circ$	$\Psi / ^\circ$	$\omega / ^\circ$
$\alpha$ -L-Rhap-(1->3)- $\alpha$ -D-Glcp	36 (34)	-36 (166)	-	39 (42)	-29 (155)	-
$\beta$ -D-Glcp-(1->6)- $\alpha$ -D-Glcp	(0)	-176	-78 (-60)	(42)	126 (145)	166 (170)
$\beta$ -D-Glcp-(1->4)- $\alpha$ -D-Glcp	54 (47)	-6 (3)	-	49 (39)	-11 (-19)	-
$\alpha$ -D-Glcp-(1->3)- $\alpha$ -L-Rhap	-36	-41	-	-26 (-13)	-19 (-35)	-
$\alpha$ -L-Rhap-(1->3)- $\beta$ -D-Glcp	31	-24	-	34 (45)	-24 (-39)	-

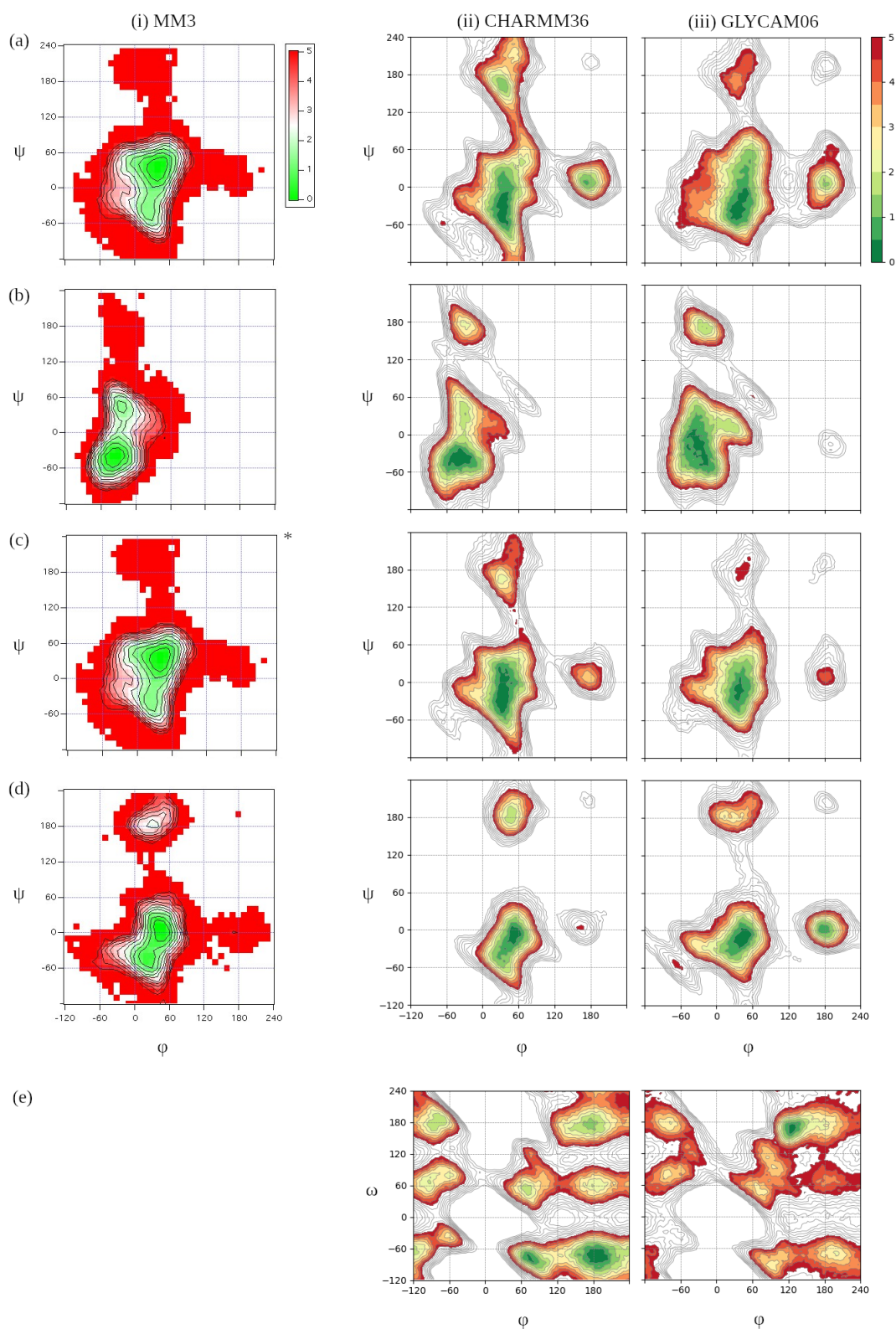
**Table A3:** Dihedral values calculated from Metadynamics simulation of substituted O25a and O25b disaccharides in vacuum. (Tables A1 and A2 contain unsubstituted bond values for these RUs)

Glycosidic Bonds	CHARMM36 minima		
	$\varphi / ^\circ$	$\Psi / ^\circ$	$\omega / ^\circ$
$\beta$ -D-GlcpNAc(1->4)- $\alpha$ -D-Glcp	56	-1	-
$\alpha$ -D-Glcp-(1->3)- $\alpha$ -L-FucpNAc	-39	21	-
$\alpha$ -L-FucNAc-(1->3)- $\beta$ -D-GlcpNAc	51	26	-
$\alpha$ -D-Glcp-(1->3)- $\alpha$ -L-RhapAc	-29	11	-
$\alpha$ -L-RhapAc-(1->3)- $\beta$ -D-GlcpNAc	51	-4	-

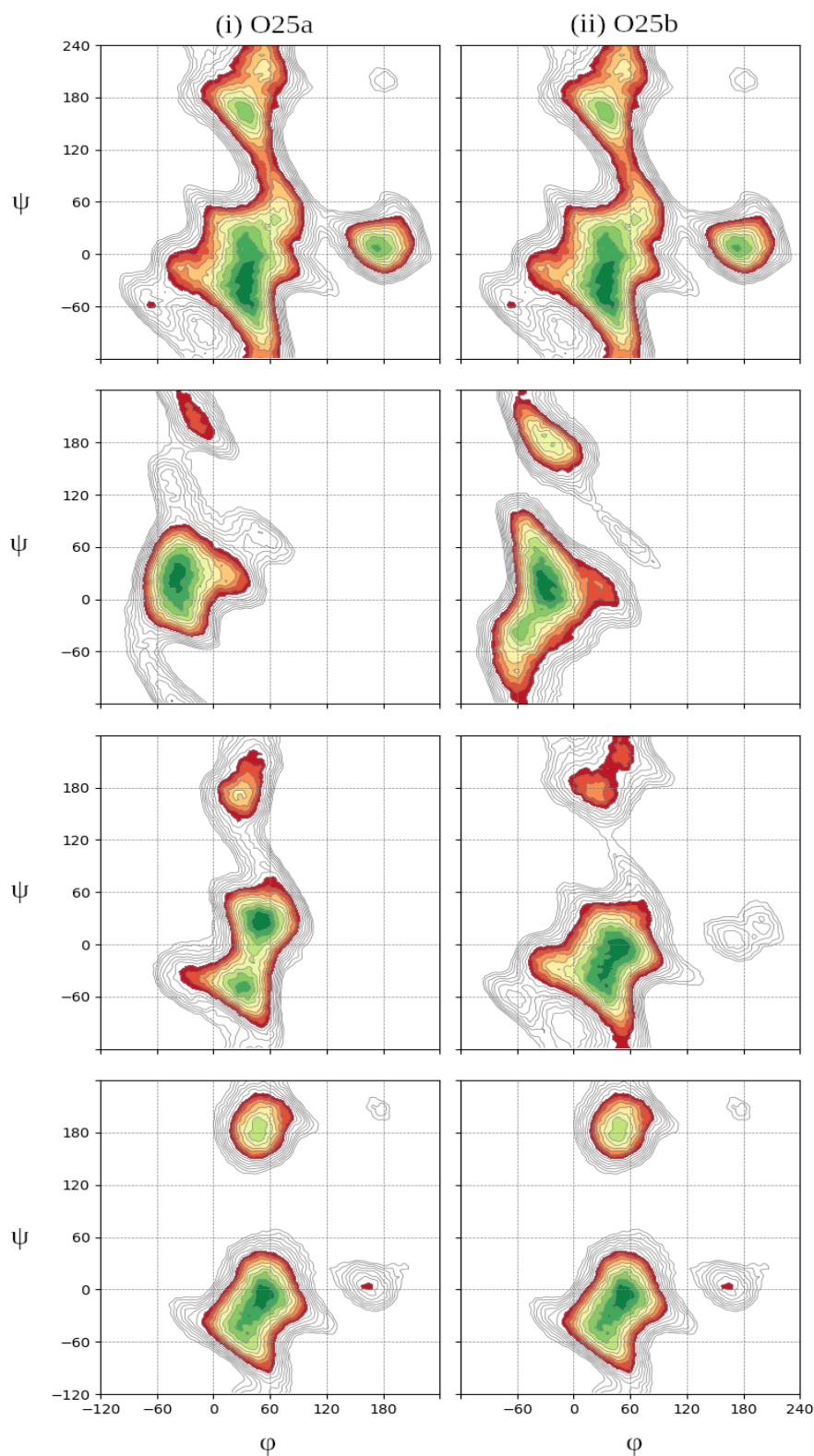


**Figure A1:** PMF plots of (i) MM3 (Böhm et. al., 1996<sup>52</sup>), (ii) CHARMM36 and (iii) GLYCAM06 data for representative disaccharides of the RU of O25a. Contour increments are at 0.5 kcal/mol up to a cutoff of 9 kcal/mol. Regions below 5 kcal/mol are highlighted. MM3 plots are of data from 10 ns of MD simulation. Bonds: (a) aR13aG (b) aG13aF (c) aF13bG (d) bG14aG (e) bG16aG.





**Figure A2:** PMF plots of (i) MM3 (Böhm et. al., 1996<sup>52</sup>), (ii) CHARMM36 and (iii) GLYCAM06 data for representative disaccharides of the O25b RU. Contour increments are at 0.5 kcal/mol up to a cutoff of 9 kcal/mol. Regions below 5 kcal/mol are highlighted. MM3 plots are of data from 10 ns of MD simulation. Bonds: (a) aR13aG (b) aG13aR (c) aR13bG (d) bG14aG (e) bG16aG.



**Figure A3:** PMF plots of representative disaccharides of the (i) O25a and (ii) O25b RUs. Contour increments are at 0.5 kcal/mol up to a cutoff of 9 kcal/mol. Regions below 5 kcal/mol are highlighted. Bonds: (i)(a) aR13aG, (i)(b) aG13aFNAc, (i)(c) aFNAc13bGNAc, (i)(d) bGNAc14aG; (ii)(a) aR13aG, (ii)(b) aG13aRAc, (ii)(c) aRAc13bGNAc, (ii)(d) bGNAc14aG.



**Appendix B: Example NAMD configuration file for metadynamics simulations**

```

#Input
    structure      .././.././models/CHARMM/vacuum/initial_dihedrals/aLRha13aDGlc.psf
    coordinates    .././.././models/CHARMM/vacuum/initial_dihedrals/aLRha13aDGlc.pdb
    set temperature 310
    set outputname  aLRha13aDGlc_100ns_md
    firsttimestep  0
    paraTypeCharmm on
    parameters     .././.././force_field_parameters/par_all36_carb_altered_ribitol.prm
    temperature     $temperature

# Force-Field Parameters
    exclude        scaled1-4
    1-4scaling      1.0
    cutoff          12.0
    switching       on
    switchdist      10.0
    pairlistdist    14.0

# Integrator Parameters
    timestep        1.0
    rigidBonds      all
    nonbondedFreq   1
    fullElectFrequency 2
    stepspercycle   10

# Constant Temperature Control
    langevin        on
    langevinDamping 1
    langevinTemp     $temperature
    langevinHydrogen off

# Pressure Control for nPT Ensemble
    useFlexibleCell no
    langevinPiston   off

# Output
    outputName      $outputname
    restartfreq     500
    dcdfreq         1000
    xstFreq         1000
    outputEnergies  1000
    outputPressure  1000

#Metadynamics
    colvars         on
    colvarsConfig    .././.././simulations/CHARMM/vacuum/colvars/aLRha13aDGlc_md.in

# Minimization
    minimize        10000
    reinitvels      $temperature

#Execution
    run             100000000

```

**Appendix C: Example metadynamics colvars definition file for disaccharide fragments**

```

colvarsTrajFrequency 500
colvarsRestartFrequency 1000
colvar {
    name Phi
    lowerBoundary -180.0
    upperBoundary 180.0
    width 2.5
    dihedral {group1 {
        atomnumbers {25}
    }
    group2 {
        atomnumbers {24}
    }
    group3 {
        atomnumbers {14}
    }
    group4 {
        atomnumbers {12}
    }
    }
}
colvar {
    name Psi
    lowerBoundary -180.0
    upperBoundary 180.0
    width 2.5
    dihedral {
        group1 {
            atomnumbers {24}
        }
        group2 {
            atomnumbers {14}
        }
        group3 {
            atomnumbers {12}
        }
        group4 {
            atomnumbers {13}
        }
    }
}
metadynamics {
    name aLRha13aDGlc_dihedrals
    colvars Phi Psi
    hillWeight 0.05
    newHillFrequency 500
    dumpFreeEnergyFile yes
    writeHillsTrajectory on
    hillwidth 2.5
}

```

## Appendix D: Example Gnuplot script for vacuum PMF surfaces

```
reset
cd '<DIRECTORY>'
set title '<TITLE> minimum at (<COLVARX>,<COLVARY>)'
set contour base
set cntrparam level incremental 0, 1, 9
unset surface
set table $dataset
splot '<PMF>'
unset table
set xrange[-180:180]
set yrange[-180:180]
set xlabel 'phi'
set ylabel 'psi'
set ytic (-120, -60, 0, 60, 120)
set xtic (-120, -60, 0, 60, 120)
set grid front
set label at <COLVARX>, <COLVARY> " " point pointtype 3 pointsize 0.75 lc rgb "blue"
unset key
unset colorbox
set terminal svg
set term svg size 400, 350
set output '<OUTPUT>'
plot $dataset w l lt -1
```

**Appendix E: Example NAMD configuration file for solution phase simulations**

```

# Input
    structure                .././.././models/CHARMM/solvated/updated_dihedrals/CSU_o25a_3RU.psf
    coordinates              .././.././models/CHARMM/solvated/updated_dihedrals/CSU_o25a_3RU.pdb
    set temperature          300
    set outputname           CSU_o25a_3RU_200ns
    firsttimestep            0
    paraTypeCharmm           on
    parameters               .././.././force_field_parameters/par_all36_carb_altered_ribitol.prm
    temperature              $temperature

# Force-Field Parameters
    exclude                  scaled1-4
    1-4scaling                1.0
    COMmotion                no
    dielectric                1.0
    switching                 on
    switchdist                12.0
    cutoff                    15.0
    pairlistdist              18.0

# Integrator Parameters
    timestep                  1.0
    nonbondedFreq            1
    fullElectFrequency        2
    stepspercycle             10

# Constant Temperature Control
    langevin                  on
    langevinDamping           1

# Periodic Boundary Conditions
    cellBasisVector1          70. 0. 0.
    cellBasisVector2          0. 70. 0.
    cellBasisVector3          0. 0. 70.
    cellOrigin                 0.0 0.1 0.0
    wrapAll                    on
    PME                       yes
    PMEGridSizeX              70
    PMEGridSizeY              70
    PMEGridSizeZ              70

# Constant Pressure Control (variable volume)
    useGroupPressure          yes
    useFlexibleCell           no

```

```
useConstantArea      no
langevinPiston        on
langevinPistonTarget  1.01325
langevinPistonPeriod  100.
langevinPistonDecay    50.
langevinPistonTemp    $temperature
```

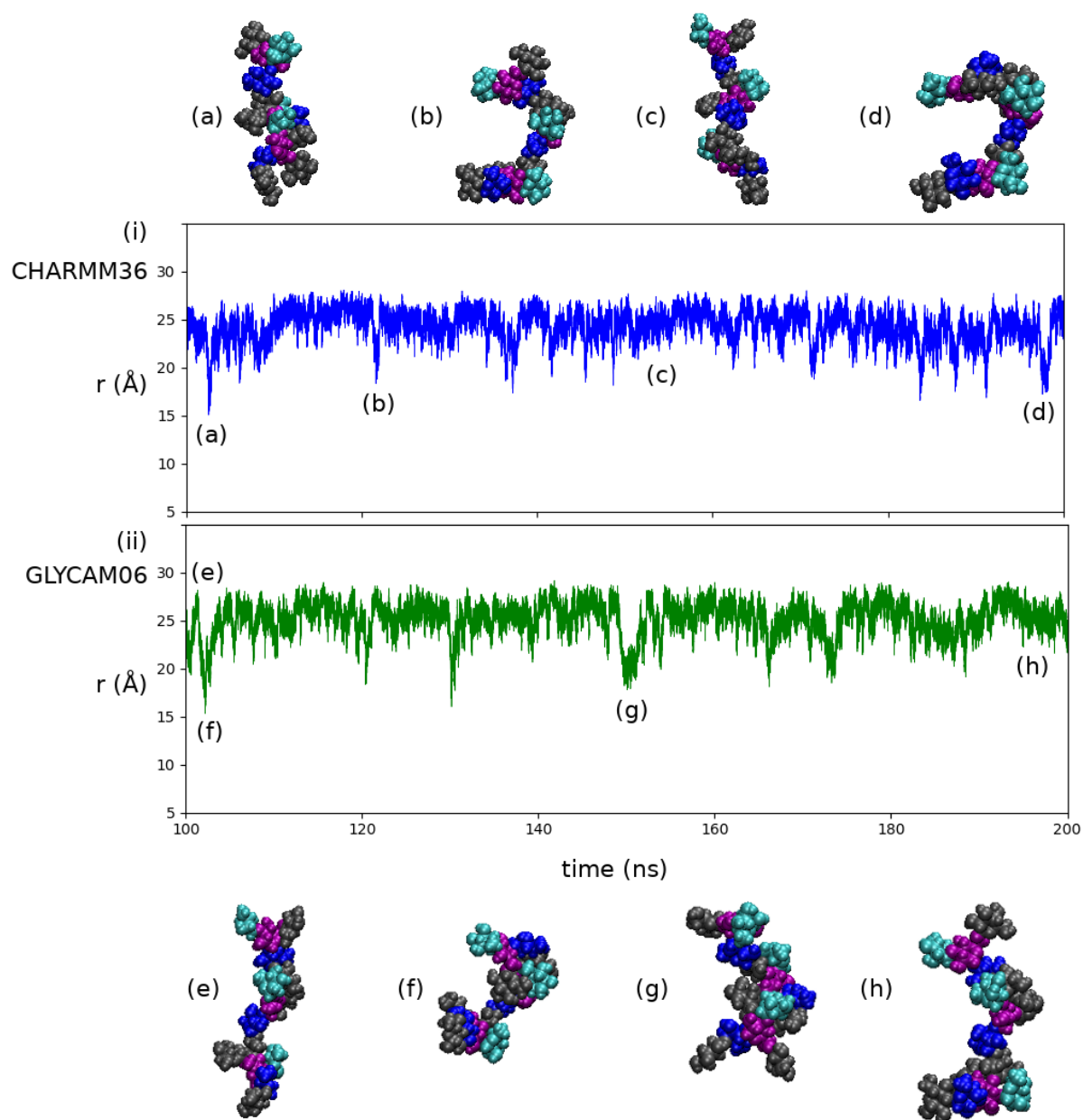
## # Output

```
outputName            $outputname
restartfreq            10000
dcdfreq               250
xstFreq               2500
outputEnergies         250
outputPressure         500
```

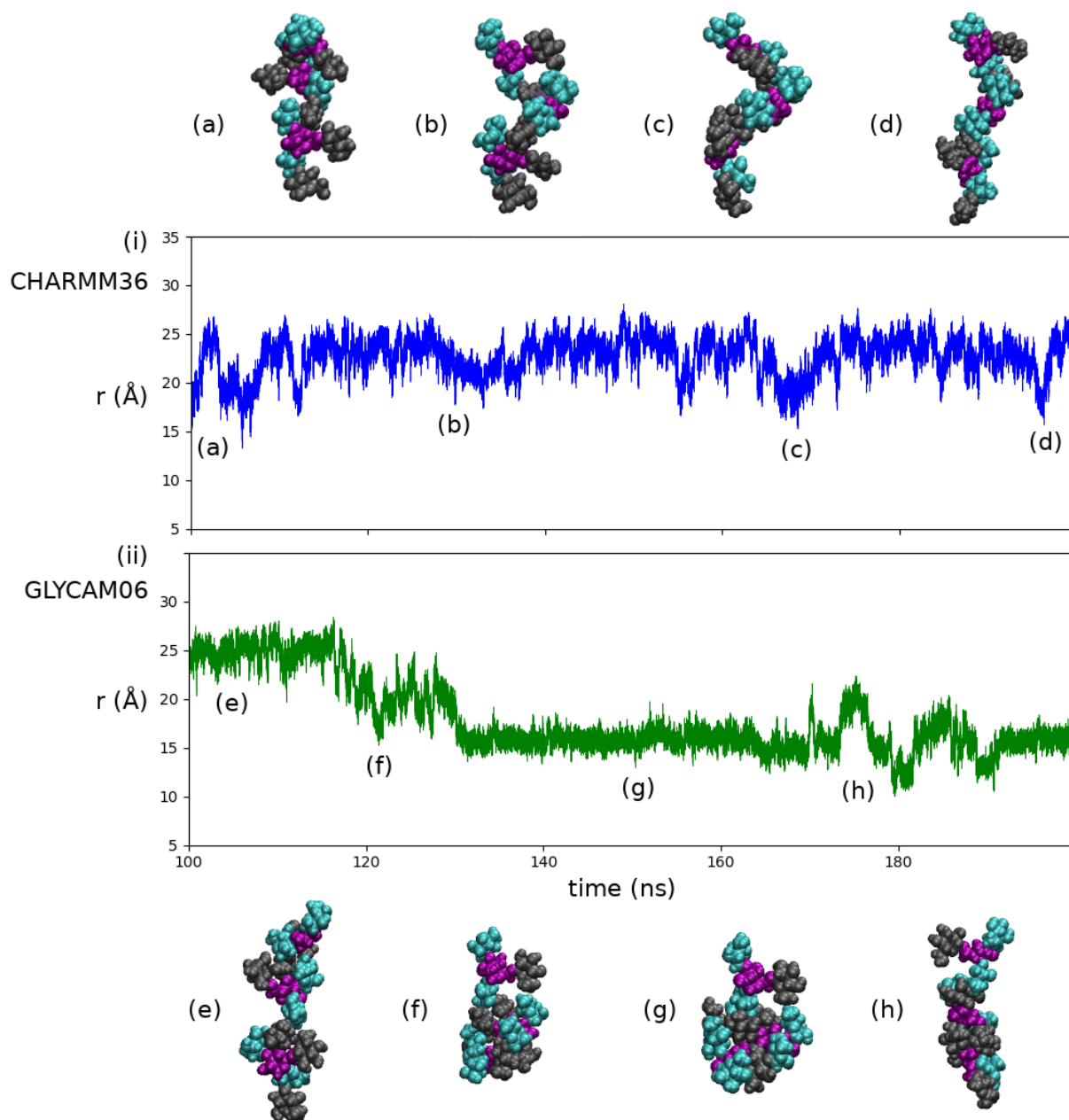
## # Execution

```
for {set i 10} {$i <= 300} {incr i 5} {
    set temperature $i
    langevinTemp $temperature
    minimize 500
    run 8000
}
firsttimestep 0
run 200000000
```

# Appendix F: Time series plot of $r$ of O25a' and O25b' with selected conformations using the CHARMM36 and GLYCAM06 force fields



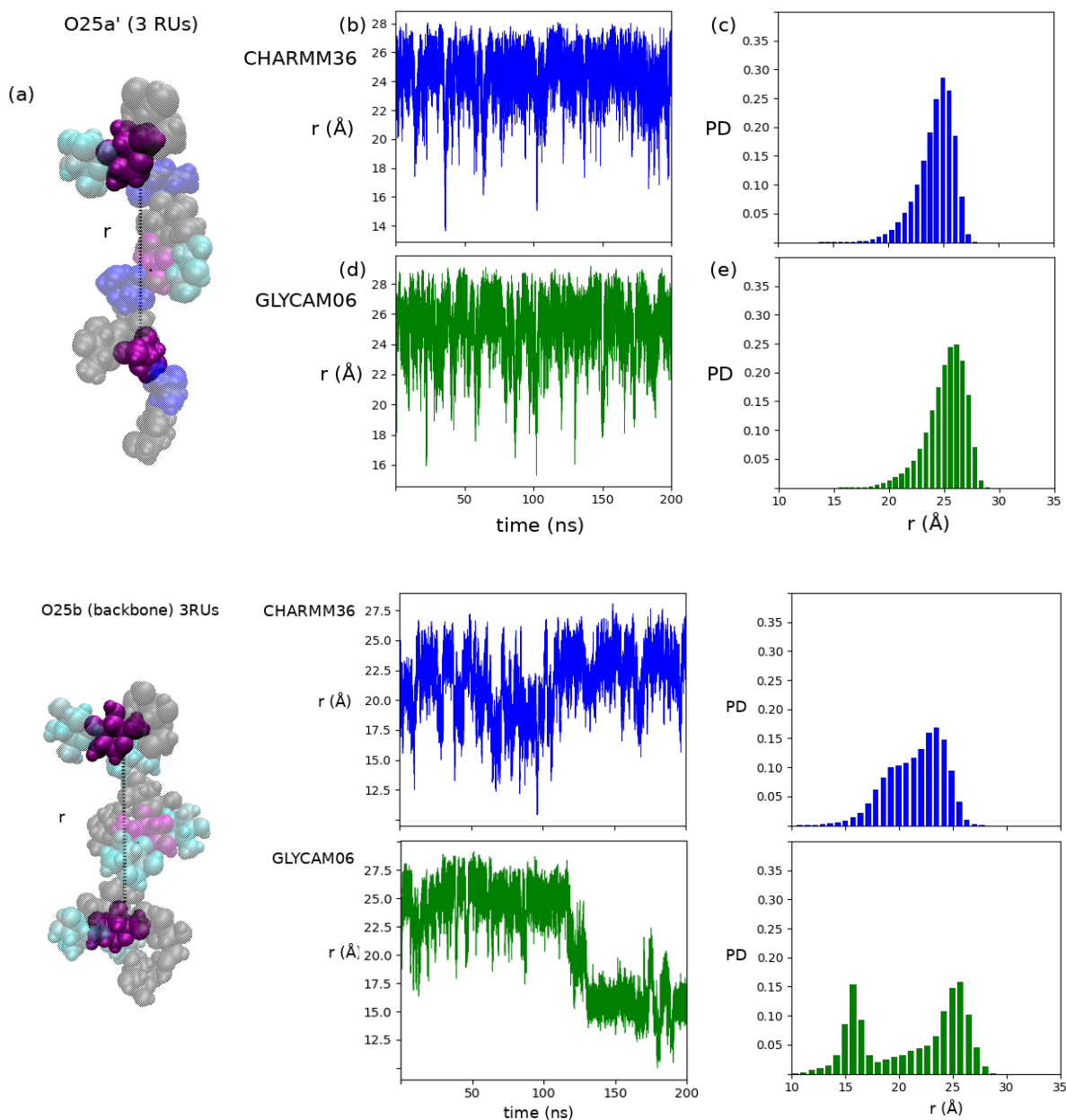
*Figure F1: Time series plots from the simulation of a 3 RU O25a' oligosaccharide extension using the (i) CHARMM36 and (ii) GLYCAM06 force fields. Points (a), (b), (c), (d), (e), (f), (g) and (h) are snapshots of conformers of interest extracted from the CHARMM36 or GLYCAM06 trajectories at the indicated simulation times.  $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc in grey,  $\alpha$ -L-Rha in cyan and  $\alpha$ -L-Fuc in blue.*



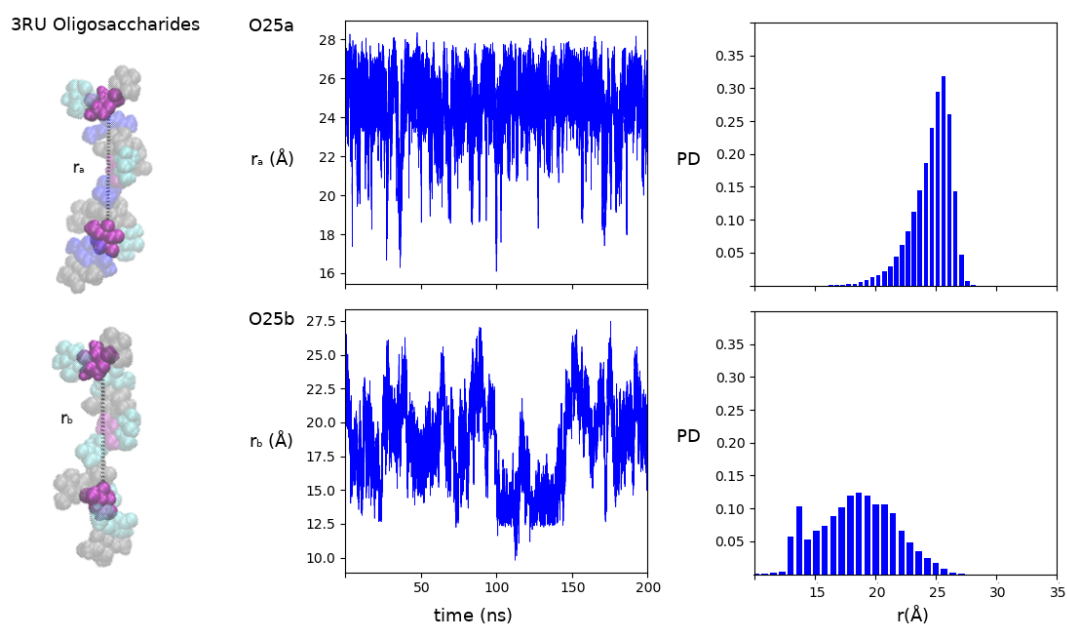
**Figure F2:** Time series plots from the simulation of a 3 RU O25b' oligosaccharide extension using the (i) CHARMM36 and (ii) GLYCAM06 force fields. Points (a), (b), (c), (d), (e), (f), (g) and (h) are snapshots of conformers of interest extracted from the CHARMM36 or GLYCAM06 trajectories at the indicated simulation times.  $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc in grey,  $\alpha$ -L-Rha in cyan.



**Appendix G: Time series and probability density plots of  $r$  of O25a', O25b', O25a and O25b for 200 ns of simulation with CHARMM36 and GLYCAM06**



**Figure G1:** (a) Model of O25b' showing the selected measure of end-to-end extension,  $r$ .  $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc in grey,  $\alpha$ -L-Rha in cyan. (left) Time series end-to-end distance plots of (b) O25b' using CHARMM36 (d) O25b' using GLYCAM06. (x), (y) and (z) represent the positions of conformers illustrated in Figure 11. (right) Probability density plots for the end-to-end distance of (c) O25b' using CHARMM36 and (e) O25b' using GLYCAM06.



**Figure G2:** (a) Model of O25b' showing the selected measure of end-to-end extension,  $r$ .  $\alpha$ -D-Glc in purple,  $\beta$ -D-Glc in grey,  $\alpha$ -L-Rha in cyan. (left) Time series end-to-end distance plots of (b) O25b' using CHARMM36 (d) O25b' using GLYCAM06. (x), (y) and (z) represent the positions of conformers illustrated in Figure 11. (right) Probability density plots for the end-to-end distance of (c) O25b' using CHARMM36 and (e) O25b' using GLYCAM06.

## Appendix H: Extract of VMD Hydrogen bond analysis of O25b

Cutoff values of 3.5Å and 30 degrees were applied in VMD.

donor	acceptor	occupancy
<b>BGLC14-Side-O6</b>	<b>BGLC11-Main-O</b>	<b>45.47%</b>
ARHM7-Side-O4	BGLC14-Side-C6	2.07%
<b>ARHM7-Side-O4</b>	<b>BGLC14-Side-O6</b>	<b>13.93%</b>
<b>BGLC6-Side-O4</b>	<b>ARHM7-Side-O5</b>	<b>9.25%</b>
ARHM10-Side-C5	AGLC8-Side-O4	22.66%
<b>AGLC8-Side-O2</b>	<b>ARHM7-Side-OA</b>	<b>17.11%</b>
<b>AGLC3-Side-O2</b>	<b>ARHM2-Side-OA</b>	<b>32.31%</b>
<b>BGLC9-Side-O6</b>	<b>BGLC6-Main-O</b>	<b>49.28%</b>
AGLC13-Side-C6	BGLC11-Main-O	4.11%
BGLC14-Side-O6	BGLC11-Main-C	8.80%
AGLC8-Side-C5	ARHM7-Side-O4	2.96%
BGLC6-Side-O4	ARHM7-Side-C5	4.10%
BGLC11-Side-C6	ARHM10-Side-O4	1.13%
BGLC9-Side-C6	ARHM5-Side-O5	1.39%
ARHM10-Side-C5	BGLC11-Side-O5	28.57%
ARHM15-Side-C5	AGLC13-Side-O4	3.75%
BGLC9-Side-O6	BGLC6-Main-C	9.59%
ARHM5-Side-C5	AGLC3-Side-O4	19.36%
ARHM5-Side-C5	BGLC6-Side-O5	30.82%
ARHM2-Side-C5	BGLC1-Side-O4	1.36%
BGLC11-Side-CT	ARHM12-Side-OA	1.26%
BGLC11-Side-C5	BGLC9-Side-O2	1.80%
ARHM10-Side-C3	BGLC11-Side-O5	8.04%
BGLC11-Side-O4	ARHM12-Side-C5	5.06%
<b>BGLC11-Side-O4</b>	<b>ARHM12-Side-O5</b>	<b>13.10%</b>
ARHM5-Side-C3	BGLC6-Side-O5	5.90%
ARHM2-Side-O4	AGLC3-Side-C5	5.03%
ARHM2-Side-O4	AGLC3-Side-O5	7.61%
<b>ARHM2-Side-O4</b>	<b>BGLC9-Side-O6</b>	<b>18.10%</b>
AGLC13-Side-O4	ARHM15-Side-O5	8.73%
AGLC8-Side-C6	BGLC6-Main-O	3.22%
ARHM12-Side-C2	BGLC11-Main-O	0.24%
BGLC11-Side-CT	BGLC14-Side-O5	5.48%
BGLC11-Side-CT	AGLC13-Side-O5	2.78%
BGLC6-Side-CT	ARHM7-Side-OA	2.88%
AGLC13-Side-C5	ARHM12-Side-O4	2.39%
AGLC13-Side-O4	ARHM15-Side-C5	4.66%
AGLC3-Side-C5	ARHM2-Side-O4	3.85%
ARHM2-Side-O4	BGLC9-Side-C6	2.42%
<b>BGLC11-Side-O6</b>	<b>ARHM10-Side-O4</b>	<b>10.15%</b>
<b>ARHM5-Side-O4</b>	<b>BGLC6-Side-O6</b>	<b>13.17%</b>
BGLC1-Side-O4	ARHM2-Side-C5	8.09%
<b>BGLC1-Side-O4</b>	<b>ARHM2-Side-O5</b>	<b>14.82%</b>
<b>AGLC13-Side-O2</b>	<b>ARHM12-Side-OA</b>	<b>13.20%</b>
<b>BGLC6-Side-O4</b>	<b>BGLC14-Side-O4</b>	<b>10.81%</b>
<b>BGLC14-Side-O4</b>	<b>BGLC6-Side-O4</b>	<b>10.49%</b>
BGLC6-Side-O6	BGLC14-Side-O3	5.24%
<b>ARHM10-Side-O4</b>	<b>BGLC11-Side-O6</b>	<b>12.20%</b>

**Appendix I: Extract of VMD Hydrogen bond analysis of O25a**

Cutoff values of 3.5Å and 30 degrees were applied in VMD. Only polar atoms were assessed.

donor	acceptor	occupancy
<b>BGLC11-Side-O4</b>	<b>AFUC12-Side-O5</b>	<b>14.82%</b>
AFUC7-Main-N	AGLC8-Side-O2	4.19%
AGLC13-Side-O4	ARHM15-Side-O5	8.52%
AGLC3-Side-O2	AFUC2-Main-N	0.06%
BGLC6-Side-O6	AGLC3-Side-O3	0.19%
<b>BGLC1-Side-O4</b>	<b>AFUC2-Side-O5</b>	<b>22.09%</b>
AGLC3-Side-O2	AFUC2-Main-O	2.48%
AFUC7-Side-O4	AGLC8-Side-O5	0.64%
AFUC7-Main-N	BGLC6-Main-O	1.04%
ARHM5-Side-O4	BGLC6-Side-O5	0.09%
AFUC12-Main-N	AGLC13-Side-O2	5.45%
AFUC2-Main-N	AGLC3-Side-O2	2.72%
AFUC2-Main-N	BGLC1-Main-O	0.45%
<b>BGLC6-Side-O4</b>	<b>AFUC7-Side-O5</b>	<b>15.43%</b>
AFUC12-Main-N	BGLC11-Main-O	1.32%
BGLC11-Side-O6	ARHM10-Side-O4	5.25%
ARHM10-Side-O4	BGLC11-Side-O6	5.47%
AGLC13-Side-O2	AFUC12-Main-O	4.50%
AGLC8-Side-O2	AFUC7-Main-O	4.30%
BGLC6-Side-O6	ARHM5-Side-O4	3.37%
ARHM5-Side-O4	BGLC6-Side-O6	3.66%